

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:42 ; Search time 46.4649 Seconds  
(without alignments)  
3958.655 Million cell updates/sec

Title: US-09-989-981A-6  
Perfect score: 3326  
Sequence: 1 MGDLSLTPGGSMGLQVNRG.....PALVILGIVVFKIRDHLISR 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	3326	100.0	651	5	AAU96984	Aau96984 Human ABC
2	3326	100.0	651	5	AAE13290	Aae13290 Human sit
3	3326	100.0	651	6	AAE31704	Aae31704 Human ABC
4	3323	99.9	651	5	AAU96992	Aau96992 Human ABC
5	3321	99.8	651	5	AAU96990	Aau96990 Human ABC
6	3321	99.8	651	5	AAU96989	Aau96989 Human ABC
7	3319	99.8	651	5	AAU96993	Aau96993 Human ABC
8	3304	99.3	649	5	ABP52128	Abp52128 Homo sapi
9	2748.5	82.6	652	5	AAE13309	Aae13309 Mouse sit

10	2744.5	82.5	652	5	AAE13289	Aae13289	Mouse	sit
11	2744.5	82.5	652	6	AAE31702	Aae31702	Mouse	ABC
12	2742.5	82.5	652	5	AAE13308	Aae13308	Mouse	sit
13	2738.5	82.3	652	5	AAU96985	Aau96985	Mouse	ABC
14	2727.5	82.0	652	5	AAU96986	Aau96986	Rat	ABCG5
15	2081	62.6	408	5	AAU96991	Aau96991	Human	ABC
16	1387.5	41.7	340	5	AAU96987	Aau96987	Hamster	A
17	1308	39.3	256	7	ADB64641	Adb64641	Human	pro
18	1234	37.1	243	5	AAU96988	Aau96988	Human	ABC
19	722	21.7	144	3	AAB41856	Aab41856	Human	ORF
20	702	21.1	672	6	AAE31703	Aae31703	Mouse	ABC
21	697	21.0	673	5	ABP52129	Abp52129	Homo sapi	
22	697	21.0	673	6	AAE31705	Aae31705	Human	ABC
23	682.5	20.5	655	5	AAU80029	Aau80029	Human	ABC
24	682.5	20.5	663	2	AAU15221	Aay15221	Breast	Ca
25	680.5	20.5	655	4	AAB60104	Aab60104	Human	tra
26	680.5	20.5	655	5	AAO14781	Aao14781	Human	BCR
27	680.5	20.5	655	5	AAU80028	Aau80028	Human	ABC
28	680.5	20.5	655	6	ABR58077	Abr58077	Human	ABC
29	680.5	20.5	655	6	ADA10917	Ada10917	Human	cDN
30	680.5	20.5	655	7	ADC54182	Adc54182	Human	bre
31	680.5	20.5	665	5	AAO14782	Aao14782	Human	BCR
32	676.5	20.3	665	5	AAO14783	Aao14783	Human	BCR
33	674.5	20.3	655	3	AAU95365	Aay95365	ATP-bindi	
34	674.5	20.3	655	4	AAU04348	Aau04348	Human	BCR
35	674.5	20.3	655	5	ABP52127	Abp52127	Homo sapi	
36	674.5	20.3	655	5	ABB07270	Abb07270	Human	BCR
37	674.5	20.3	655	6	ABU63376	Abu63376	Human	mit
38	672.5	20.2	655	5	ABB07273	Abb07273	Human	BCR
39	660	19.8	657	5	ABB07272	Abb07272	Murine	BC
40	602.5	18.1	687	4	ABB59384	Abb59384	Drosophil	
41	600	18.0	602	4	ABB65432	Abb65432	Drosophil	
42	598.5	18.0	674	5	ABP52126	Abp52126	Homo sapi	
43	590.5	17.8	638	5	ABB98349	Abb98349	Human	ABC
44	588.5	17.7	648	3	AAG18078	Aag18078	Arabidops	
45	587.5	17.7	625	3	AAG18080	Aag18080	Arabidops	

# ALIGNMENTS

## RESULT 1

AAU96984

ID AAU96984 standard; protein; 651 AA.

XX

AC AAU96984;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 protein.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;  
KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;  
KW chromosome 2p21.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers  
FT Misc-difference 2. .15  
FT /note= "Encoded by GGTCTC"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX

PI Patel SB, Dean M;

XX

DR WPI; 2002-416483/44.

DR N-PSDB; ABK51681.

XX

PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic  
PT acid encoding the polypeptide, useful for treating sitosterolemia,  
PT arteriosclerosis and heart diseases.

XX

PS Claim 52; Page 35-36; 66pp; English.

XX

CC The present invention relates to a new mammalian ATP-binding cassette  
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
CC disease. The molecules of the invention are also useful for identifying a  
CC compound which alters ABCG5 activity level comprising contacting a cell  
CC culture or mammal which have ABCG5 polypeptide with a compound and  
CC measuring ABCG5 biological activity in the cell culture or in mammal,  
CC where an increase or decrease in ABCG5 biological activity compared to  
CC ABCG5 biological activity in a control cell culture or mammal not  
CC contacted with the compound, identifies a compound that increases or  
CC decreases ABCG5 activity respectively. The cell culture or mammal  
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
CC polypeptide in a cell culture or mammal is also compared with that of a  
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
CC Stimulation of ABCG5 activity is useful for treating or preventing  
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
CC disease. The method of the invention is useful for increasing cholesterol  
CC excretion and/or decreasing cholesterol adsorption. The present amino  
CC acid sequence represents the human ABCG5 protein of the invention. This  
CC sequence is encoded by the human ABCG5 gene located on chromosome 2p21

XX

SQ Sequence 651 AA;

Query Match 100.0%; Score 3326; DB 5; Length 651;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC 60  
|||||

Db	1	MGDLSSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Qy	61	RQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Qy	121	RREQFQDCFSYVLQSDTLLSSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLDQPKVMLFDEPTTGLDCMTANQIVVLLEVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLDQPKVMLFDEPTTGLDCMTANQIVVLLEVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVATM	480
Qy	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Db	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVVFKIRDHLISR	651

# RESULT 2

AAE13290

ID AAE13290 standard; protein; 651 AA.

XX

AC AAE13290;

XX

DT 12-FEB-2002 (first entry)

XX

DE Human sitosterolaemia susceptibility gene (SSG) protein.

XX

KW Human; sitosterolaemia susceptibility gene; SSG; atherosclerosis;  
KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; therapy;  
KW gall stone; coronary heart disease; cardiovascular disease; arthritis;  
KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 2p21.



XX  
 OS Homo sapiens.  
 XX  
 PN WO200179272-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PF 18-APR-2001; 2001WO-US012758.  
 XX  
 PR 18-APR-2000; 2000US-0198465P.  
 PR 15-MAY-2000; 2000US-0204234P.  
 XX  
 PA (TULA-) TULARIK INC.  
 XX  
 PI Tian H, Schultz J, Shan B;  
 XX  
 DR WPI; 2002-017598/02.  
 DR N-PSDB; AAD22009.  
 XX  
 PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,  
 PT useful for screening a compound that increases the level of expression or  
 PT activity of SSG polypeptide for treating sterol-related disorder.  
 XX  
 PS Claim 19; Fig 8; 105pp; English.  
 XX  
 CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene  
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)  
 CC binding cassette (ABC) family cholesterol transporter. SSG is useful for  
 CC identifying a compound useful in the treatment or prevention of a sterol-  
 CC related disorder, including sitosterolaemia, hyperlipidaemia,  
 CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or  
 CC nutritional deficiencies. SSG is also useful for treating cholesterol-  
 CC associated diseases or conditions including coronary heart disease and  
 CC other cardiovascular diseases, and sitosterolaemia-associated condition  
 CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG  
 CC expression cassette is useful in the production of transgenic non-human  
 CC animals. SSG genes and their homologues are useful as tools for a number  
 CC of applications including diagnosing sitosterolaemia and other  
 CC cardiovascular disorders, for forensics and paternity determinations, and  
 CC for treating any of a large number of SSG associated diseases. The  
 CC present sequence is human SSG protein. Human SSG is located on chromosome  
 CC 2p21  
 XX  
 SQ Sequence 651 AA;

Query Match 100.0%; Score 3326; DB 5; Length 651;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWW	DITSC 60
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWW	DITSC 60
Qy	61	RQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120

Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMGLFLLFFVLRVRSNVLKGAIQDRV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVATM	480
Qy	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Db	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	651

RESULT 3

AAE31704

ID AAE31704 standard; protein; 651 AA.

XX

AC AAE31704;

XX

DT 24-MAR-2003 (first entry)

XX

DE Human ABCG5 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;  
 KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;  
 KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;  
 KW human; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;  
 KW ABCG5.

XX

OS Homo sapiens.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.  
 XX  
 PF 20-NOV-2001; 2001WO-US043823.  
 XX  
 PR 20-NOV-2000; 2000US-0252235P.  
 PR 28-NOV-2000; 2000US-0253645P.  
 XX  
 PA (TULA-) TULARIK INC.  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Hobbs HH, Shan B, Barnes R, Tian H;  
 XX  
 DR WPI; 2003-058548/05.  
 DR N-PSDB; AAD48882.  
 XX  
 PT New ABCG8 polypeptides and nucleic acids, useful for treating sterol-  
 PT related disorders e.g. sitosterolemia, hypercholesterolemia,  
 PT hyperlipidemia, gall stones, HDL deficiency, atherosclerosis, or  
 PT nutritional deficiencies.  
 XX  
 PS Claim 28; Page 78-79; 94pp; English.  
 XX  
 CC The invention relates to ATP-binding cassette (ABC) family cholesterol  
 CC transporter, ABCG8 polypeptides and polynucleotides. The invention also  
 CC provides ABCG5 polypeptides and polynucleotides. ABCG5 gene is also known  
 CC as sitosterolaemia susceptibility gene (SSG). Sequences of the invention  
 CC are useful for treating or preventing sterol-related disorders such as  
 CC sitosterolaemia, hyperlipidaemia, hypercholesterolaemia, gall stones, HDL  
 CC deficiency, atherosclerosis and nutritional deficiencies. They are also  
 CC useful in gene therapy. The present sequence is human ABCG5 protein  
 XX  
 SQ Sequence 651 AA;

Query Match 100.0%; Score 3326; DB 6; Length 651;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWW	DITSC	60
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWW	DITSC	60
Qy	61	RQQWTRQILKDVS	LYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVS	LYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRET	LHYTALLAIRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRET	LHYTALLAIRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLL	LVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLL	LVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNP	PDFD	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNP	PDFD	300

Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKS	AI	CKTLKNI	ERMKHLKTL	PMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKS	AI	CKTLKNI	ERMKHLKTL	PMVPFKTKD	360
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNL	IMGLFLLFF	VLRVRSNV	LKGAIQDRV	420	
Db	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNL	IMGLFLLFF	VLRVRSNV	LKGAIQDRV	420	
Qy	421	GLLYQFVGATPYTGMLNAVNLF	PVLR	AVSDQESQDGLYQKWQ	MMLAYALH	VL	PFSVVATM 480
Db	421	GLLYQFVGATPYTGMLNAVNLF	PVLR	AVSDQESQDGLYQKWQ	MMLAYALH	VL	PFSVVATM 480
Qy	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHL	LIGEF	TLVLLGIVQNP	NI	VNSV	VALLSIA 540
Db	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHL	LIGEF	TLVLLGIVQNP	NI	VNSV	VALLSIA 540
Qy	541	GVLVSGGFLRNIQEMPIPFKII	SYFTFQKYCSEILV	NEFYGLNFTCGSSNV	SVTTNPMC	600	
Db	541	GVLVSGGFLRNIQEMPIPFKII	SYFTFQKYCSEILV	NEFYGLNFTCGSSNV	SVTTNPMC	600	
Qy	601	AFTQGIQFIEKTCPGATSRFTMNL	LILYSFIPALVILGIV	VFKIRDHLISR	651		
Db	601	AFTQGIQFIEKTCPGATSRFTMNL	LILYSFIPALVILGIV	VFKIRDHLISR	651		

RESULT 4

AAU96992

ID AAU96992 standard; protein; 651 AA.

XX

AC AAU96992;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant E146Q protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 146

FT /note= "Wild-type Glu substituted by Gln"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX

PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PA (PATE/) PATEL S B.

PA (DEAN/) DEAN M.

XX  
 PI Patel SB, Dean M;  
 XX  
 DR WPI; 2002-416483/44.  
 XX  
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic  
 PT acid encoding the polypeptide, useful for treating sitosterolemia,  
 PT arteriosclerosis and heart diseases.  
 XX  
 PS Claim 12; Page; 66pp; English.  
 XX  
 CC The present invention relates to a new mammalian ATP-binding cassette  
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
 CC disease. The molecules of the invention are also useful for identifying a  
 CC compound which alters ABCG5 activity level comprising contacting a cell  
 CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not  
 CC contacted with the compound, identifies a compound that increases or  
 CC decreases ABCG5 activity respectively. The cell culture or mammal  
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
 CC polypeptide in a cell culture or mammal is also compared with that of a  
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
 CC Stimulation of ABCG5 activity is useful for treating or preventing  
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
 CC disease. The method of the invention is useful for increasing cholesterol  
 CC excretion and/or decreasing cholesterol adsorption. The present amino  
 CC acid sequence represents the human ABCG5 mutant E146Q protein of the  
 CC invention. Note: This sequence is not shown in the specification but is  
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
 CC 35-36 of the specification  
 XX  
 SQ Sequence 651 AA;

Query Match 99.9%; Score 3323; DB 5; Length 651;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 650; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGDLSSTLTPGSGMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWIDITSC 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 MGDLSSTLTPGSGMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWIDITSC 60  
 Qy 61 RQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 61 RQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL 120  
 Qy 121 RREQFQDCFSYVLQSDTLLSSLTVRQLTHYTALLAIRGNPGSFQKKVEAVMAELSLSHV 180  
 ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||  
 Db 121 RREQFQDCFSYVLQSDTLLSSLTVRQLTHYTALLAIRGNPGSFQKKVEAVMAELSLSHV 180  
 Qy 181 ADRLIGNYSLGGISTGERRRVSIAAQLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR 240  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 181 ADRLIGNYSLGGISTGERRRVSIAAQLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR 240

Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Qy	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Db	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	651

RESULT 5

AAU96990

ID AAU96990 standard; protein; 651 AA.

XX

AC AAU96990;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R389H protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;

KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;

KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 389

FT /note= "Wild-type Arg substituted by His"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.

XX

PF 25-SEP-2001; 2001WO-US029859.

XX

PR 25-SEP-2000; 2000US-0235268P.

XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA (PATE/) PATEL S B.  
PA (DEAN/) DEAN M.  
XX  
PI Patel SB, Dean M;  
XX  
DR WPI; 2002-416483/44.  
XX  
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic  
PT acid encoding the polypeptide, useful for treating sitosterolemia,  
PT arteriosclerosis and heart diseases.  
XX  
PS Claim 7; Page; 66pp; English.  
XX  
CC The present invention relates to a new mammalian ATP-binding cassette  
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
CC disease. The molecules of the invention are also useful for identifying a  
CC compound which alters ABCG5 activity level comprising contacting a cell  
CC culture or mammal which have ABCG5 polypeptide with a compound and  
CC measuring ABCG5 biological activity in the cell culture or in mammal,  
CC where an increase or decrease in ABCG5 biological activity compared to  
CC ABCG5 biological activity in a control cell culture or mammal not  
CC contacted with the compound, identifies a compound that increases or  
CC decreases ABCG5 activity respectively. The cell culture or mammal  
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
CC polypeptide in a cell culture or mammal is also compared with that of a  
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
CC Stimulation of ABCG5 activity is useful for treating or preventing  
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
CC disease. The method of the invention is useful for increasing cholesterol  
CC excretion and/or decreasing cholesterol adsorption. The present amino  
CC acid sequence represents the human ABCG5 mutant R389H protein of the  
CC invention. Note: This sequence is not shown in the specification but is  
CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
CC 35-36 of the specification  
XX  
SQ Sequence 651 AA;

Query Match 99.8%; Score 3321; DB 5; Length 651;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Qy	61	RQQWTRQILKDVSPLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVSPLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180

Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRRVTRNLVRNKLAVITHLLQNLMGLFLLFFVLRVRSNVLKGAIQDRV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVL PFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVL PFSVVATM	480
Qy	481	IFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNI VNSVVALLSIA	540
Db	481	IFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNI VNSVVALLSIA	540
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMFLILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMFLILYSFIPALVILGIVVFKIRDHLISR	651

RESULT 6

AAU96989

ID AAU96989 standard; protein; 651 AA.

XX

AC AAU96989;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R419H protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;  
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;  
 KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 419

FT /note= "Wild-type Arg substituted by His"

XX

PN WO200227016-A2.

XX

PD 04-APR-2002.



XX  
PF 25-SEP-2001; 2001WO-US029859.  
XX  
PR 25-SEP-2000; 2000US-0235268P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
PA (PATE/) PATEL S B.  
PA (DEAN/) DEAN M.  
XX  
PI Patel SB, Dean M;  
XX  
DR WPI; 2002-416483/44.  
XX  
PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic  
PT acid encoding the polypeptide, useful for treating sitosterolemia,  
PT arteriosclerosis and heart diseases.  
XX  
PS Claim 9; Page; 66pp; English.  
XX  
CC The present invention relates to a new mammalian ATP-binding cassette  
CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
CC disease. The molecules of the invention are also useful for identifying a  
CC compound which alters ABCG5 activity level comprising contacting a cell  
CC culture or mammal which have ABCG5 polypeptide with a compound and  
CC measuring ABCG5 biological activity in the cell culture or in mammal,  
CC where an increase or decrease in ABCG5 biological activity compared to  
CC ABCG5 biological activity in a control cell culture or mammal not  
CC contacted with the compound, identifies a compound that increases or  
CC decreases ABCG5 activity respectively. The cell culture or mammal  
CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
CC polypeptide in a cell culture or mammal is also compared with that of a  
CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
CC Stimulation of ABCG5 activity is useful for treating or preventing  
CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
CC disease. The method of the invention is useful for increasing cholesterol  
CC excretion and/or decreasing cholesterol adsorption. The present amino  
CC acid sequence represents the human ABCG5 mutant R419H protein of the  
CC invention. Note: This sequence is not shown in the specification but is  
CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
CC 35-36 of the specification  
XX  
SQ Sequence 651 AA;

Query Match 99.8%; Score 3321; DB 5; Length 651;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Qy	61	RQQWTRQILKDVSILYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVSILYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120

Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDHV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Qy	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPINVSVALLSIA	540
Db	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPINVSVALLSIA	540
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVVFKIRDHLISR	651

RESULT 7

AAU96993

ID AAU96993 standard; protein; 651 AA.

XX

AC AAU96993;

XX

DT 30-JUL-2002 (first entry)

XX

DE Human ABCG5 mutant R419P protein sequence.

XX

KW Human; ABCG5; ATP-binding cassette gene 5; sitosterolemia; cholesterol;  
 KW arteriosclerosis; heart disease; hypersterolemia; Alzheimer's disease;  
 KW mutant; mutein.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 419

FT /note= "Wild-type Arg substituted by Pro"

XX  
 PN WO200227016-A2.  
 XX  
 PD 04-APR-2002.  
 XX  
 PF 25-SEP-2001; 2001WO-US029859.  
 XX  
 PR 25-SEP-2000; 2000US-0235268P.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA (PATE/) PATEL S B.  
 PA (DEAN/) DEAN M.  
 XX  
 PI Patel SB, Dean M;  
 XX  
 DR WPI; 2002-416483/44.  
 XX  
 PT Novel mammalian ATP-binding cassette gene 5 polypeptide, and the nucleic  
 PT acid encoding the polypeptide; useful for treating sitosterolemia,  
 PT arteriosclerosis and heart diseases.  
 XX  
 PS Claim 10; Page; 66pp; English.  
 XX  
 CC The present invention relates to a new mammalian ATP-binding cassette  
 CC gene 5 (ABCG5) polypeptide. The invention is useful for identifying a  
 CC predisposition for developing sitosterolemia, arteriosclerosis or heart  
 CC disease. The molecules of the invention are also useful for identifying a  
 CC compound which alters ABCG5 activity level comprising contacting a cell  
 CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not  
 CC contacted with the compound, identifies a compound that increases or  
 CC decreases ABCG5 activity respectively. The cell culture or mammal  
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
 CC polypeptide in a cell culture or mammal is also compared with that of a  
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
 CC Stimulation of ABCG5 activity is useful for treating or preventing  
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
 CC disease. The method of the invention is useful for increasing cholesterol  
 CC excretion and/or decreasing cholesterol adsorption. The present amino  
 CC acid sequence represents the human ABCG5 mutant R419P protein of the  
 CC invention. Note: This sequence is not shown in the specification but is  
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
 CC 35-36 of the specification  
 XX  
 SQ Sequence 651 AA;

Query Match 99.8%; Score 3319; DB 5; Length 651;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 650; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC 60

Qy	61	RQQWTRQILKDVS	LYVESGQIMCILGSSSGSGKT	TLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVS	LYVESGQIMCILGSSSGSGKT	TLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRET	LHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180	
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRET	LHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180	
Qy	181	ADRLIGNYSLGGISTGERRRV	SIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240	
Db	181	ADRLIGNYSLGGISTGERRRV	SIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240	
Qy	241	RNRIVVLT	TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND	CGYPCPEHSNPFDF	300
Db	241	RNRIVVLT	TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND	CGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREI	ETSKRVQMIESAYKKS	SAICHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREI	ETSKRVQMIESAYKKS	SAICHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVL	LRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420	
Db	361	SPGVFSKLGVL	LRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDPV	420	
Qy	421	GLLYQFVGATPYTGMLNAVN	LFPVLRAVSDQESQDGLYQKWQMMLAYALHVL	PFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVN	LFPVLRAVSDQESQDGLYQKWQMMLAYALHVL	PFSVVATM	480
Qy	481	IFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEF	TLVL	LGIVQNPNI	VNSVVALLSIA 540
Db	481	IFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEF	TLVL	LGIVQNPNI	VNSVVALLSIA 540
Qy	541	GVLVSGSGFLRNIQEMPIPF	KIISYFTFQKYCSEILV	NEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVSGSGFLRNIQEMPIPF	KIISYFTFQKYCSEILV	NEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTC	PGATSRFTMNF	LILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTC	PGATSRFTMNF	LILYSFIPALVILGIVVFKIRDHLISR	651

RESULT 8

ABP52128

ID ABP52128 standard; protein; 649 AA.

XX

AC ABP52128;

XX

DT 10-OCT-2002 (first entry)

XX

DE Homo sapiens ABC transporter ABCG5 protein SEQ ID NO:80.

XX

KW ATP-binding cassette transporter; ABC transporter; modulation; D loop;

KW cancer; bacterial infection; fungal infection; protozoal infection;

KW antibacterial; fungicide; protozoacide.

XX

OS Homo sapiens.

XX

PN EP1217066-A1.  
 XX  
 PD 26-JUN-2002.  
 XX  
 PF 21-DEC-2000; 2000EP-00870316.  
 XX  
 PR 21-DEC-2000; 2000EP-00870316.  
 XX  
 PA (UYGE-) UNIV GENT.  
 XX  
 DR WPI; 2002-550404/59.  
 XX  
 PT Modulating activity of ATP-binding cassette (ABC) transporters by  
 PT influencing dimerization of nucleotide binding domains through use of D  
 PT loop sequence of an ABC transporter, or its antisense peptide or peptide  
 PT mimetic.  
 XX  
 PS Disclosure; Fig 3; 290pp; English.  
 XX  
 CC The present invention describes a method (M1) for modulating the activity  
 CC of ATP-binding cassette (ABC) transporters by influencing the  
 CC dimerisation of the nucleotide binding domains comprises using: (a) a  
 CC polypeptide (polyP) consisting of 5-50 amino acids comprising the D loop  
 CC sequence of an ABC transporter (ABP52049 to ABP52091); (b) a polyP  
 CC consisting of the D loop sequence of an ABC transporter; (c) a peptide  
 CC mimetic or antisense peptide of (a) or (b). ABC transporters have  
 CC antibacterial, fungicide and protozoacide activities. (M1) is useful for  
 CC selectively modulating the activity of ABC transporters belonging to the  
 CC group of multidrug transporter/P-glycoproteins. Bacterial, fungal or  
 CC protozoal ABC transporters are involved in the infection of a mammal or  
 CC in the induction of resistance to antibiotics or drugs in a mammal. (M1)  
 CC is useful for preventing, treating or alleviating diseases associated  
 CC with functionality of an ABC transporter. ABP52092 to ABP52140 represent  
 CC ABC transporter proteins given in the exemplification of the present  
 CC invention  
 XX  
 SQ Sequence 649 AA;

Query Match 99.3%; Score 3304; DB 5; Length 649;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 649; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Qy	61	RQQWTRQILKDVSPLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVSPLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLDPKVMLF--PTTGLDCMTANQIVVLLVELAR	238

Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	239	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	298
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKS AICHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	299	YMDLTSVDTQSKEREIETSKRVQMIESAYKKS AICHKTLKNIERMKHLKTLPMVPFKTKD	358
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	359	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	418
Qy	421	GLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQDGLYQKWQMMLAYALHVL PFSVVATM	480
Db	419	GLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQDGLYQKWQMMLAYALHVL PFSVVATM	478
Qy	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNI VNSVVALLSIA	540
Db	479	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNI VNSVVALLSIA	538
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	539	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	598
Qy	601	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	651
Db	599	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	649

RESULT 9

AAE13309

ID AAE13309 standard; protein; 652 AA.

XX

AC AAE13309;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein variant #2.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis; mutein;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; mutant;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.

XX

OS Mus sp.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 28

FT /note= "Wild type Gly substituted with Ala"

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX  
PR 18-APR-2000; 2000US-0198465P.  
PR 15-MAY-2000; 2000US-0204234P.  
XX  
PA (TULA-) TULARIK INC.  
XX  
PI Tian H, Schultz J, Shan B;  
XX  
DR WPI; 2002-017598/02.  
XX  
PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,  
PT useful for screening a compound that increases the level of expression or  
PT activity of SSG polypeptide for treating sterol-related disorder.  
XX  
PS Disclosure; Page; 105pp; English.  
XX  
CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene  
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)  
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for  
CC identifying a compound useful in the treatment or prevention of a sterol-  
CC related disorder, including sitosterolaemia, hyperlipidaemia,  
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or  
CC nutritional deficiencies. SSG is also useful for treating cholesterol-  
CC associated diseases or conditions including coronary heart disease and  
CC other cardiovascular diseases, and sitosterolaemia-associated condition  
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG  
CC expression cassette is useful in the production of transgenic non-human  
CC animals. SSG genes and their homologues are useful as tools for a number  
CC of applications including diagnosing sitosterolaemia and other  
CC cardiovascular disorders, for forensics and paternity determinations, and  
CC for treating any of a large number of SSG associated diseases. The  
CC present sequence is mouse SSG protein variant obtained by replacing Gly28  
CC with Ala. Note: The present sequence is not shown in the specification  
CC but is derived from mouse SSG protein referred as SEQ ID NO: 1 (AAE13289)  
CC and shown in figure 7 of the specification  
XX  
SQ Sequence 652 AA;

Qy 240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPFD 299  
 ||:||||:||||| |||||:||||| ||| |||:||||| |||||  
 Db 241 RRDRIVIVTTHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNNCGYPCEHSNPFD 300  
 Qy 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK 359  
 |||||:||||| |||||:| |:|:| | || |:||| |:||||| |||||  
 Db 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360  
 Qy 360 DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419  
 | ||:| |||||:|||| ||| ||:|||||:|:|:|:| ||||:|||  
 Db 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNLTKGAVQDR 420  
 Qy 420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVAT 479  
 ||||| |||||:||||| |||||:||||| |||||:||| |||||:|||  
 Db 421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480  
 Qy 480 MIFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSVALLSI 539  
 :|||||:||||| |||||:||||| |||||:||||| |||||:|||||  
 Db 481 VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSIVALLSI 540  
 Qy 540 AGVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599  
 |:|:|:|:|:|:| ||: ||||| ||||| ||||| ||| |: :||  
 Db 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGSNTSMLNHMP 600  
 Qy 600 CAFTQGIQFIEKTCPGATSRTMNFILILYSFIPALVILGIVVFKIRDHLISR 651  
 || |||:||||| ||||| |||||:|:|:|:| |||||  
 Db 601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYLISR 652

RESULT 10

AAE13289

ID AAE13289 standard; protein; 652 AA.

XX

AC AAE13289;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis;

KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia;

KW gall stone; coronary heart disease; cardiovascular disease; arthritis;

KW xanthoma; haemolytic anaemia; transgenic animal; chromosome 17; therapy.

XX

OS Mus sp.

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.



XX  
PI Tian H, Schultz J, Shan B;  
XX  
DR WPI; 2002-017598/02.  
DR N-PSDB; AAD22008.  
XX  
PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,  
PT useful for screening a compound that increases the level of expression or  
PT activity of SSG polypeptide for treating sterol-related disorder.  
XX  
PS Claim 19; Fig 7; 105pp; English.  
XX  
CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene  
CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)  
CC binding cassette (ABC) family cholesterol transporter. SSG is useful for  
CC identifying a compound useful in the treatment or prevention of a sterol-  
CC related disorder, including sitosterolaemia, hyperlipidaemia,  
CC hypercholesterolaemia, gall stones, HDL deficiency, atherosclerosis or  
CC nutritional deficiencies. SSG is also useful for treating cholesterol-  
CC associated diseases or conditions including coronary heart disease and  
CC other cardiovascular diseases, and sitosterolaemia-associated condition  
CC including arthritis, xanthomas and chronic haemolytic anaemia. SSG  
CC expression cassette is useful in the production of transgenic non-human  
CC animals. SSG genes and their homologues are useful as tools for a number  
CC of applications including diagnosing sitosterolaemia and other  
CC cardiovascular disorders, for forensics and paternity determinations, and  
CC for treating any of a large number of SSG associated diseases. The  
CC present sequence is mouse SSG protein. Mouse SSG is located on chromosome  
CC 17  
XX  
SQ Sequence 652 AA;

```

Db      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPVFFKTK 360
Qy      360 DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419
      | ||:| |||||:| || ||:| || ||:| || ||:| || ||:| || ||:| ||
Db      361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAQDR 420
Qy      420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVAT 479
      ||||| |||||:| ||||| ||||:| || |||||:| ||
Db      421 VGLLYQLVGATPYTGMLNAVNLFPMRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSVIAT 480
Qy      480 MIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSI 539
      :|||:| |||||:| ||||| |||||:| ||||| |||||:| |||||
Db      481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSIVALLSI 540
Qy      540 AGVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599
      :|:|:| ||||| ||:| ||||| ||||| ||||| || ||:| ||
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
Qy      600 CAFTQGIQFIEKTCPGATSRTMNFILILYSFIPALVILGIVVFKIRDHLISR 651
      || |||||:| ||||| ||||| ||||| ||||| ||||| |||||
Db      601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYILISR 652

```

RESULT 11

AAE31702

ID AAE31702 standard; protein; 652 AA.

XX

AC AAE31702;

XX

DT 24-MAR-2003 (first entry)

XX

DE Mouse ABCG5 protein.

XX

KW ABC family cholesterol transporter; ABCG8; sterol-related disorder;  
KW sitosterolaemia; hyperlipidaemia; hypercholesterolaemia; gall stone;  
KW HDL deficiency; atherosclerosis; nutritional deficiency; gene therapy;  
KW mouse; ATP-binding cassette; sitosterolaemia susceptibility gene; SSG;  
KW ABCG5.

XX

OS Mus sp.

XX

PN WO200281691-A2.

XX

PD 17-OCT-2002.

XX

PF 20-NOV-2001; 2001WO-US043823.

XX

PR 20-NOV-2000; 2000US-0252235P.

PR

28-NOV-2000; 2000US-0253645P.

XX

PA (TULA-) TULARIK INC.

PA (TEXA ) UNIV TEXAS SYSTEM.

XX

PI Hobbs HH, Shan B, Barnes R, Tian H;

XX

DR WPI; 2003-058548/05.

Qy	1	MGDLSSLTPGSSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS	59
Db	1	MGELPFLSPEGARGPHINRGSLSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS	60
Qy	60	CRQQWTRQILKDVSILYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGDEVYVNGRA	119
Db	61	CQQKWDRQILKDVSILYIESGQIMCILGSSSGSGKTTLLDAISGRLRRRTGTLERGEVFNVC	120
Qy	120	LRREQFQDCFSYVLQSDTLLSSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH	179
Db	121	LRRDQFQDCFSYVLQSDVFLSSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSH	180
Qy	180	VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGLDCMTANQIVVLLVELA	239
Db	181	VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Qy	240	RRNRIVVLTIIHQPRSELFQLEFDKIAILSFGEILFCGTPAEMLDFFNDCCGYPCPEHSNPF	299
Db	241	RRDRIVIVTIIHQPRSELFQHFQDKIAILTYGELVFCGTPPEMLGFFNNCCGYPCPEHSNPF	300
Qy	300	FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFCK	359
Db	301	FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFCK	360
Qy	360	DSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLQLNIMGLFLLFFVLRVRSNVLKGAIQDR	419
Db	361	DPPGMFGKLGVLRLRRVTRNLVRNKLAVIMRLVQNLIMGLFLIFYLLRVQNNLTGKAVQDR	420
Qy	420	VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVAT	479
Db	421	VGLLYQLVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT	480
Qy	480	MIFSSVCYWTGLGHPHVARFYGYSAAALLAPHLIGEFLLTVLLGIVQNPINIVSNVALLSI	539

```

      :|||||:|||||:|||||:|||||
Db      481 VIFSSVCYWTGLGYPEVARFGYFSAALLAPHLIGFLLVLLGIVQNPNIIVNSIVALLSI 540

Qy      540 AGVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599
      :|:|:||||:||||| ||: ||||| ||||| ||||| || | : ||
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600

Qy      600 CAFTQGIQFIEKTCPGATSRTMNFILILYSFIPALVILGIVVFKIRDHLISR 651
      || ||:||||| ||||| ||||| ||||| ||:|:|:|
Db      601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYISR 652

```

RESULT 12

AAE13308

ID AAE13308 standard; protein; 652 AA.

XX

AC AAE13308;

XX

DT 12-FEB-2002 (first entry)

XX

DE Mouse sitosterolaemia susceptibility gene (SSG) protein variant #1.

XX

KW Mouse; sitosterolaemia susceptibility gene; SSG; atherosclerosis; mutein;  
 KW sterol-related disorder; hyperlipidaemia; hypercholesterolaemia; mutant;  
 KW gall stone; coronary heart disease; cardiovascular disease; arthritis;  
 KW xanthoma; haemolytic anaemia; transgenic animal; therapy; variant.

XX

OS Mus sp.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 17

FT /note= "Wild type Ile substituted with Leu"

XX

PN WO200179272-A2.

XX

PD 25-OCT-2001.

XX

PF 18-APR-2001; 2001WO-US012758.

XX

PR 18-APR-2000; 2000US-0198465P.

PR 15-MAY-2000; 2000US-0204234P.

XX

PA (TULA-) TULARIK INC.

XX

PI Tian H, Schultz J, Shan B;

XX

DR WPI; 2002-017598/02.

XX

PT Novel sitosterolemia susceptibility gene polypeptide and polynucleotide,  
 PT useful for screening a compound that increases the level of expression or  
 PT activity of SSG polypeptide for treating sterol-related disorder.

XX

PS Disclosure; Page; 105pp; English.

XX

CC The invention relates to an isolated Sitosterolaemia Susceptibility Gene  
 CC (SSG) polypeptide. SSG is a member of adenosine triphosphate (ATP)





CC disease. The molecules of the invention are also useful for identifying a  
 CC compound which alters ABCG5 activity level comprising contacting a cell  
 CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not  
 CC contacted with the compound, identifies a compound that increases or  
 CC decreases ABCG5 activity respectively. The cell culture or mammal  
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
 CC polypeptide in a cell culture or mammal is also compared with that of a  
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
 CC Stimulation of ABCG5 activity is useful for treating or preventing  
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
 CC disease. The method of the invention is useful for increasing cholesterol  
 CC excretion and/or decreasing cholesterol adsorption. The present amino  
 XX acid sequence represents the mouse ABCG5 protein of the invention  
 SQ Sequence 652 AA;

Query Match 82.3%; Score 2738.5; DB 5; Length 652;  
 Best Local Similarity 80.1%; Pred. No. 2.7e-279;  
 Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;

Qy 1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59  
 ||:| |:| |:| :||| ||| | | |||:| ||||:| |||:| |  
 Db 1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60

Qy 60 CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119  
 |:|:| |||||:|||||:|||||:|||| | | |||:| |  
 Db 61 CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFNVC 120

Qy 120 LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH 179  
 |||:||||| ||||| ||||| |||:| | : : ||||| |||||  
 Db 121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTLSL 180

Qy 180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLDEPTTGLDCMTANQIVVLLVELA 239  
 |||:|:|:| |||:||||| |||||: ||||| |||||:| |||  
 Db 181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240

Qy 240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPF 299  
 ||:|||:| ||||| |||||:|:|:| ||| || | |||:| |||||  
 Db 241 RRDRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNNGYPCPEHSNPF 300

Qy 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK 359  
 ||||| |||||:| ||||| |||||:| |:|:| | | |:| ||| :| ||||| |||||  
 Db 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPVPFKTK 360

Qy 360 DSPGVFESKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419  
 | ||:| ||||| |||||:| ||| || |:| ||||| |||:|:|:| |||:| |  
 Db 361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNTLKGAVQDR 420

Qy 420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSV 479  
 ||||| ||||| |||||:| ||||| ||||| ||||:| | |||||:| |  
 Db 421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPSVI 480

Qy 480 MIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSVALLSI 539





measuring ABCG5 biological activity in the cell culture or in mammal, where an increase or decrease in ABCG5 biological activity compared to ABCG5 biological activity in a control cell culture or mammal not contacted with the compound, identifies a compound that increases or decreases ABCG5 activity respectively. The cell culture or mammal comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The ABCG5 biological activity, or level of ABCG5 mRNA, or level of the polypeptide in a cell culture or mammal is also compared with that of a second cell culture or mammal comprising a wild type ABCG5 polypeptide. Stimulation of ABCG5 activity is useful for treating or preventing hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's disease. The method of the invention is useful for increasing cholesterol excretion and/or decreasing cholesterol adsorption. The present amino acid sequence represents the rat ABCG5 protein of the invention. (Updated on 07-AUG-2003 to correct OS field.)

SQ Sequence 652 AA;

Query Match 82.0%; Score 2727.5; DB 5; Length 652;  
Best Local Similarity 79.4%; Pred. No. 3.9e-278;  
Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

Qy	1	MGDLSLTPGGSGMGLQVNRGSSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDDITS	59
Db	1	MGELPFLSPEGARGPHNNRGSSQSSLEEGSVTGSEARHSLGVLNVSFVSNSNRVGPWWNIKS	60
Qy	60	CRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGGEVYVNGRA	119
Db	61	CQQKWDRKILKDVSLYIESGQTMCILGSSSGSGKTTLLDAISGRLLRRTGTLEGEVVFVNGCE	120
Qy	120	LRREQFQDCFSYVLQSDTLLSSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH	179
Db	121	LRRDQFQDCVSYLLQSDVFLSSSLTVRETLRYTAMLALRSSSADFYDKKVEAVLTELSLSH	180
Qy	180	VADRLIGNYSLGGISSTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELA	239
Db	181	VADQMIGNYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANHIVLLVELA	240
Qy	240	RRNRIVVLTIHQPRSELFQLEFDKIAILSFGEILFCGTPAEMLDFFNDCGYPCPEHSNPF	299
Db	241	RRNRIVIVTIIHQPRSELFHHFDKIAILTYGELVFCGTPPEMLGFFNCCGYPCPEHSNPF	300
Qy	300	FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPF	359
Db	301	FYMDLTSVDTQSREREIETYKRVQMLESAFRQSDICHKILENIERTRHKLTLPMVPF	360
Qy	360	DSPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR	419
Db	361	NPPGMFCKLGVLLRRVTRNLMRNKQVVIMRLVQNLIMGLFLIFYLLRVQNMLKGAVQDR	420
Qy	420	VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLFPFSV	479
Db	421	VGLLYQLVGATPYTGMLNAVNLFPMRLRAVSDQESQDGLYQKWQMLLAYVLHALPFSIV	480
Qy	480	MIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNIIVNSVVAL	539
Db	481	VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLLTVLLGMVQNPNIIVNSIVALL	540



CC compound which alters ABCG5 activity level comprising contacting a cell  
 CC culture or mammal which have ABCG5 polypeptide with a compound and  
 CC measuring ABCG5 biological activity in the cell culture or in mammal,  
 CC where an increase or decrease in ABCG5 biological activity compared to  
 CC ABCG5 biological activity in a control cell culture or mammal not  
 CC contacted with the compound, identifies a compound that increases or  
 CC decreases ABCG5 activity respectively. The cell culture or mammal  
 CC comprises a mutated ABCG5 polypeptide or a wild type polypeptide. The  
 CC ABCG5 biological activity, or level of ABCG5 mRNA, or level of the  
 CC polypeptide in a cell culture or mammal is also compared with that of a  
 CC second cell culture or mammal comprising a wild type ABCG5 polypeptide.  
 CC Stimulation of ABCG5 activity is useful for treating or preventing  
 CC hypersterolemia, arteriosclerosis, heart disease and/or Alzheimer's  
 CC disease. The method of the invention is useful for increasing cholesterol  
 CC excretion and/or decreasing cholesterol adsorption. The present amino  
 CC acid sequence represents the human ABCG5 mutant R408X protein of the  
 CC invention. Note: This sequence is not shown in the specification but is  
 CC derived from the wild-type human ABCG5 protein (AAU96984) given on pages  
 CC 35-36 of the specification  
 XX  
 SQ Sequence 408 AA;

Query Match 62.6%; Score 2081; DB 5; Length 408;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-210;  
 Matches 408; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Qy	61	RQQWTRQILKDVSPLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVSPLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVR	408
Db	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVR	408

Search completed: February 27, 2004, 06:44:21  
 Job time : 48.4649 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:11:48 ; Search time 14.7508 Seconds  
(without alignments)  
2278.426 Million cell updates/sec

Title: US-09-989-981A-6  
Perfect score: 3326  
Sequence: 1 MGDLSSTPGGSMGLQVNRG.....PALVILGIVVFKIRDHLISR 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	682.5	20.5	655	4	US-09-245-808-1	Sequence 1, Appli
2	674.5	20.3	655	4	US-09-767-594-1	Sequence 1, Appli
3	436.5	13.1	1296	4	US-09-614-912-140	Sequence 140, App
4	373.5	11.2	617	4	US-09-614-912-138	Sequence 138, App
5	334.5	10.1	539	4	US-09-614-912-144	Sequence 144, App
6	263.5	7.9	653	4	US-09-543-681A-5411	Sequence 5411, Ap
7	259	7.8	384	4	US-09-489-039A-9127	Sequence 9127, Ap
8	258.5	7.8	210	4	US-09-543-681A-8215	Sequence 8215, Ap
9	253.5	7.6	373	4	US-09-543-681A-7638	Sequence 7638, Ap
10	249.5	7.5	245	4	US-09-540-236-3618	Sequence 3618, Ap
11	245	7.4	344	4	US-09-489-039A-13987	Sequence 13987, A

12	244	7.3	248	4	US-09-134-001C-3731	Sequence 3731, Ap
13	244	7.3	1280	2	US-08-752-447-2	Sequence 2, Appli
14	244	7.3	1280	4	US-09-316-167-2	Sequence 2, Appli
15	244	7.3	1280	4	US-09-397-233-2	Sequence 2, Appli
16	243.5	7.3	276	4	US-09-489-039A-13021	Sequence 13021, A
17	240	7.2	1279	2	US-08-784-649A-2	Sequence 2, Appli
18	240	7.2	1279	4	US-09-672-810-6	Sequence 6, Appli
19	240	7.2	1280	2	US-08-583-276-19	Sequence 19, Appl
20	240	7.2	1280	4	US-09-767-594-2	Sequence 2, Appli
21	240	7.2	1280	4	US-09-672-810-5	Sequence 5, Appli
22	240	7.2	1280	6	5206352-4	Patent No. 5206352
23	239.5	7.2	358	4	US-09-489-039A-7399	Sequence 7399, Ap
24	239.5	7.2	1684	3	US-08-665-259-25	Sequence 25, Appl
25	239.5	7.2	1684	3	US-08-762-500-25	Sequence 25, Appl
26	239.5	7.2	1704	3	US-08-762-500-75	Sequence 75, Appl
27	239	7.2	229	4	US-09-134-000C-3584	Sequence 3584, Ap
28	237	7.1	254	4	US-09-489-039A-13102	Sequence 13102, A
29	236.5	7.1	329	4	US-09-107-532A-4844	Sequence 4844, Ap
30	234.5	7.1	266	4	US-09-252-991A-26488	Sequence 26488, A
31	232.5	7.0	286	4	US-09-540-236-2931	Sequence 2931, Ap
32	232.5	7.0	360	4	US-09-543-681A-6466	Sequence 6466, Ap
33	230.5	6.9	242	4	US-09-134-001C-3832	Sequence 3832, Ap
34	230.5	6.9	280	4	US-09-489-039A-12451	Sequence 12451, A
35	229.5	6.9	672	4	US-09-489-039A-12101	Sequence 12101, A
36	229	6.9	460	4	US-09-134-001C-3369	Sequence 3369, Ap
37	229	6.9	1307	1	US-08-395-246C-2	Sequence 2, Appli
38	228.5	6.9	261	4	US-09-252-991A-33060	Sequence 33060, A
39	228.5	6.9	589	4	US-09-328-352-7592	Sequence 7592, Ap
40	228.5	6.9	922	4	US-09-489-039A-8938	Sequence 8938, Ap
41	228	6.9	243	4	US-09-489-039A-13621	Sequence 13621, A
42	227	6.8	1280	4	US-09-672-810-7	Sequence 7, Appli
43	226.5	6.8	250	4	US-09-328-352-7153	Sequence 7153, Ap
44	226	6.8	243	4	US-09-543-681A-5911	Sequence 5911, Ap
45	225.5	6.8	347	4	US-09-543-681A-5961	Sequence 5961, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-245-808-1

; Sequence 1, Application US/09245808

; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; TITLE OF INVENTION: encodes it

; FILE REFERENCE: Ross Umb conversion

; CURRENT APPLICATION NUMBER: US/09/245,808

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Human MCF-7/AdrVp cells  
US-09-245-808-1

Query Match 20.5%; Score 682.5; DB 4; Length 655;  
Best Local Similarity 29.2%; Pred. No. 2.9e-64;  
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;

Qy	21	SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE	77
		:         : :   : : : :     :   :   : : :	
Db	13	SQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK	72
Qy	78	SGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT	137
		:       :       :     :   :   :   :   :   :	
Db	73	PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLING-APRPAKFKNCSGYVQDDV	129
Qy	138	LLSSLTVRETLHYTALLAIRGNPG-SFQKKVEAVMAELSLSHVADRLLIGNYSLGGISTG	196
		: :           : :   : : : :   :         :   :   :	
Db	130	VMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIRGVSGG	189
Qy	197	ERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVLLVELARRNRIVVLTIHQPRSEL	256
		:     : :               : :   : : :   : :       :	
Db	190	ERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRITIFSIFHQPYSI	249
Qy	257	FQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQ-----SK	312
		:     : : :   :       :       :     :   : :   : : :	
Db	250	FKLFDSLTLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR	309
Qy	313	ERE-----IETSKR----VQMIESAYKKSACHKT-----LKNIERMKHLKTLPMVPF	356
		:   :     : : :   : :   :   :   :   : : :   : : :	
Db	310	EEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY	369
Qy	357	KTKDSPGVFSLKGLVLLRRVTRNLVRNKLAVITRLLQNLMGLFL--LFFVLRVRSNVLKG	414
		: : :   :   :   : : :   : : :   : :   :   :	
Db	370	TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKNDST----	421
Qy	415	AIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRAVSDQESQDGLYQKWQMLLAYAL-HVLP	473
		:   : :   : : :       : :   :   :   :   :	
Db	422	GIQNRAGVLF-FLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP	480
Qy	474	FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPINVNSV	533
		: : : :   : :       :   : : : : : :   : : :	
Db	481	MTMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVVSV	537
Qy	534	VALLSIAGV--LVGSGFLRNIQEMPIPKIISYFTFQKYCSEILVNEFYGLNFTCGSSN	591
		: :   : :       : : : :   : :	
Db	538	TLLMTICFVMMIFSGLLVNLTIIASWLSWLQYFSIPRYGFTALQHNEFLGQNFPCG---	594
Qy	592	VSVTTNPMCAFTQGIQFIEKTCPG	615
		: :       :	
Db	595	LNATGNNPCNYA-----TCTG	610

RESULT 2  
US-09-767-594-1  
; Sequence 1, Application US/09767594

```

; Patent No. 6521635
; GENERAL INFORMATION:
; APPLICANT: Bates, Susan
; APPLICANT: Robey, Robert
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: Inhibition of MXR Transport by Acridine Derivatives
; FILE REFERENCE: 015280-402100US
; CURRENT APPLICATION NUMBER: US/09/767,594
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 60/177,410
; PRIOR FILING DATE: 2000-01-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human mitoxanthrone resistance (MXR)/BRCP/ABCP
; OTHER INFORMATION: protein
US-09-767-594-1

```

```

Query Match          20.3%; Score 674.5; DB 4; Length 655;
Best Local Similarity 29.0%; Pred. No. 2.2e-63;
Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

```

```

Qy      21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSPLYVE 77
      || : | ||| : : | : : :: :||: ||: ::|| :: :
Db      13 SQGNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK 72

Qy      78 SGQIMCILGSSSGSGKTTLLDAMSGR LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT 137
      | : ||| :| ||::||| :: | :| :| :|| | | | : ||:| |
Db      73 PG-LNAILGPTGGGKSSLLDVL AARKDPSG-LSGDVLING-APRPANFKCNSGYVVQDDV 129

Qy     138 LLSSLTVRET LHYTALLAIRRGNGP-SFQKKVEAVMAELSLSHVADRLIGNYS LGGISTG 196
      :: :||| | :| | : : : | : || | ||| :| : | :| |
Db     130 VMGTLTVREN LQFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQFIRGVSGG 189

Qy     197 ERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSEL 256
      ||:| || :| : || :: ||||| ||| ::|| :::: | : : ||||| :
Db     190 ERKRTSIGMELITDPSILSLDEPTTGLDSSTANAVLLLLKRM SKQGRTIIFS IHQPRYSI 249

Qy     257 FQLFDKIAILSFGELIFCGTPAEMLDFFND CGYPCPEHSNPFDFYMDLTSVDTQ----SK 312
      |:||| : :| : | :| | | :| || | :|| ||::| : : | : ::
Db     250 FKLFD SLTLLASGR LMFHGP AQEALGYFESAGYHCEAYNNPADFFLDIINGDSTAV ALNR 309

Qy     313 ERE-----IETSKR----VQMIESAYKKS AICHKT-----LKNIERMKHLKTLPMVPF 356
      | : || ||| : : : | | : :| | | :| : : :
Db     310 EEDFKATEIIEPSKQDKPLIEKLA EIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY 369

Qy     357 KTKDSPGVFSKLGVL LRRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFVLRVRSNVLKG 414
      | :| : :| :||: | | | :: : :|| : :| | : |
Db     370 TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDST---- 421

Qy     415 AIQDRVGLLYQFVGATPYTGMLNAVNLF PVLRAVSDQESQDGLYQKWQMMLAYAL-HVLP 473

```

```

      ||:| |:|: |:      :|| || |: : | | |: | | :||
Db      422 GIQNRAGVLF-FLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVS SYFLGKLLSDLLP 480
Qy      474 FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNI VNSV 533
      :::||: : |: ||| |: | | | : : : | | :||:
Db      481 MRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVV SVA 537
Qy      534 VALLSIAGV--LVGSGFLRNIQEMPIPFKII SYFTFQKYCSEILVVNEFYGLNFTCGSSN 591
      ||:| | :| || |: : : | |: :| | || | || |
Db      538 TLLMTICFVEMMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF CPG--- 594
Qy      592 VSVTTNPMCAFTQGIQFIEKTC PG 615
      :: | | | : | | |
Db      595 LNATGNNPCNYA-----TCTG 610

```

RESULT 3

US-09-614-912-140

; Sequence 140, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

```

; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 140
; LENGTH: 1296
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-614-912-140

```

```

Query Match          13.1%; Score 436.5; DB 4; Length 1296;
Best Local Similarity 27.6%; Pred. No. 3.7e-37;

```



Matches 173; Conservative 99; Mismatches 245; Indels 109; Gaps 27;

```

Qy      84 ILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLT 143
      :||  ||||| |::| :  ||| || |  :|:| : :|
Db      9  LLGPPSSGKTTLLLAGKLDPSLRGGEVTYNGFELEEFVAQKTAAYISQTDVHVGEMT 68

Qy     144 VRETLHYTAL-----LAIRRGNPG-----SFQK--KVEAVMAELSLSHV 180
      |:||| :|  || |  |  |  :| | : | :
Db      69 VKETLDFSARCQGVGTYKDLLTELARREKEAGIRPEPEVDLFMKATSMEGVESSLQTDYT 128

Qy     181 -----ADRLIGNYSLGGISTGERRRVSIQAQLLQDPKVMLFDEPTTGLDCMTANQI 231
      || :|:  ||| |::||:  ::  ||: || :||| | ||
Db     129 LRILGLDICADTIVGDQMQRGISGGQKKRVTTGEMIVGPTKVLFMDEISTGLDSSTTFQI 188

Qy     232 VVLLVELARRNRIVVL-TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND CGYP 290
      | | ::  :| :: || | |:| | | :|| |::: |  :|:| ||:
Db     189 VKCIQQIVHLGEATILMSLLQAPETFELFDDIILLSEGQIVYQGPREYVLEFFESCGR 248

Qy     291 CPEHSNPFDFYMDLTS-----VDTQSKEREIETSKRVQMIESAYKKSACHKTLKNI 342
      |||  || :||  |  | | | : |  :|: :  :
Db     249 CPERKGTADFLQEVTSKKDQEQYWADKHPYRYSISVSEFAQ----RFKRFHV-----GL 298

Qy     343 ERMKHLKTLPMVFF-KTKDSPG--VFSKLGVLRLRVTRN-----LVRNKLAVITRL 391
      :  ||  ||| ||:  ||| |  : :  : ||  | : :
Db     299 QLENHLS----VPFDKTRSHQAALVFSKQSVSTTELLKASFKEWLLIKRNSFYIYFKTI 354

Qy     392 QNLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSDQ 451
      | :|: |  || : :  : | | :| :|| : : :| | | |  :
Db     355 QLIIVALVASTVFLRTQMHTRN--LDD--GFVY--IGALLFSLIVNMFNGFAELSLTITR 408

Qy     452 -----ESQDGL-YQKWQMMLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVARFGYFSAA 505
      : :| | | |  |  : :||:: ::: | |:| | || | |
Db     409 LPVFFKHRDLLFYPAWIFTLPNVILRIPFSIIESIVWVIVTYTYTIGFAPEADRF--FKQL 466

Qy     506 LLAPHLIGEF---LTLVLLGIVQNPNIIVNSVVALLSIAGVLVGSGFLRNIQEMPIPFKII 562
      ||  || :  |  |: : : | : || :  :| |||  :| | |
Db     467 LLV-FLIQQMAGGLFRATAGLCRSMIIAQTGALALLIFFVLG-GFL-----LPKAF-IP 518

Qy     563 SYFTFQKYCSEI-----LVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGA 616
      :: : : | :  | |||| :  | | |  || :|  ||
Db     519 KWWIWGYWVSPLMYGYNALAVNEFYSPRW----MNKFVLDNNGVPKRLGIALME----GA 570

Qy     617 TSRETMNFIILYSFIPALVILGIVVF 642
      | :  :| | :|| :|
Db     571 NIFTDKNWF----WIGAAGLLGFTMF 592

```

RESULT 4

US-09-614-912-138

; Sequence 138, Application US/09614912

; Patent No. 6677502

; GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Rafalski, Antoni

; APPLICANT: Orozco, Buddy

; APPLICANT: Miao, Gou-Hau

; APPLICANT: Famodu, Omolayo O.

```

; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Caimi, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 138
; LENGTH: 617
; TYPE: PRT
; ORGANISM: Zea mays
US-09-614-912-138

```

```

Query Match          11.2%; Score 373.5; DB 4; Length 617;
Best Local Similarity 25.2%; Pred. No. 6.5e-31;
Matches 140; Conservative 101; Mismatches 205; Indels 109; Gaps 23;

```

```

Qy      67 QILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFL-GEVYVNGRALRREQF 125
      |:|::|:      | : ::| ||:|||||:| ::|| : | :: |:: : | : |
Db      37 QLLREVTGSFRPGVLTALMGVSGAGKTTLMDVLAGR--KTGGYIEGDIRIAGYPKNQATF 94

Qy     126 QDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSF-----QKKVEAVMAEL 175
      | |::|      :||||:| |::| | :      ||      : |:: || :
Db      95 ARISGYCEQNDIHSPPQVTVRESLIYSAFLRL----PGKIGDQEITDDIKMQFVDEVMELV 150

Qy     176 SLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLL 235
      | :: | |::| : |::| :|::|::|| :|: :| :: |||:|||| | :: :
Db     151 ELDNLRDALVGLPGITGLSTEQRKRLTIAVELVANPSIIFMDEPTSGLDARAAAIVMRTV 210

Qy     236 VELARRNRIVVLTIHQPRSELFQLFDKIAILS-FGELIFCG----TPAEMLDFFNDG-GY 289
      | || ||||| :|:| |::: | |::|:| |::|:| |:::| |
Db     211 RNTVDTGRTVVCTIHQPSIDIFESFDELLLLKRGGQVIYSGKLGRNSQKMVEYFEAIPGV 270

Qy     290 P-CPEHSNPFDFYMDLTSVDTQSK-----EREIETSKRVQMIESAYKKSAICHKTLKNIE 343
      | : || : ::::|| |: :      : |||      ||:: : | |
Db     271 PKIKDKYNPATWMLEVSSVATEVRLKMDFAKYYETS-----DLYKQNKVLVNQLSQPE 323

Qy     344 RMKHLKTLPMVP-----FKTKDSPGVFSKLGVL-----RRVTRNLVRNKLAVITR 389
      |      | |: | : |      |      | ||| : :
Db     324 -----PGTSDLYFPTEYSQSTIGQFKACLWKQWLTYWRSPDYNLVRYSFTLLVA 372

```

Qy 390 LLQNLIMGLFLLFFVLRVRSNVLKGAIQDR-----VGLLYQFVGATPYTGMLNAVNLF 443  
 || | |: |: |: |: |: |: |: |  
 Db 373 LLLGSIF-----WRIGTN-----MEDATTLGMVIGAMYT---AVMFIGINNCSTVQP 416  
 Qy 444 VL---RAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARF- 499  
 |: | |: |: |: |: |: |: |: |  
 Db 417 VVSIERTV FYRERAAGMYSAMPYAIAQVVIEIPYVFVQTYYTTLIVYAMMSFQWTAVKFF 476  
 Qy 500 -----GYFSAALLAPHLIGEFLLVLLGIVQ---NPN-IVNSVVALLSIAGVLVGSGFLR 550  
 ||| || |: |: |: |: |: |: |  
 Db 477 WFFFISYFS-----FLYFTYYGMMAVSISPNEVASIFAAAFSLFNLFSGFF- 524  
 Qy 551 NIQEMPIPFKIISYF 565  
 | || |: |  
 Db 525 -IPRPRIPGWWIWYY 538

RESULT 5

US-09-614-912-144  
 ; Sequence 144, Application US/09614912  
 ; Patent No. 6677502  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Allen, Steve  
 ; APPLICANT: Rafalski, Antoni  
 ; APPLICANT: Orozco, Buddy  
 ; APPLICANT: Miao, Gou-Hau  
 ; APPLICANT: Famodu, Omolayo O.  
 ; APPLICANT: Lee, Jian Ming  
 ; APPLICANT: Sakai, Hajime  
 ; APPLICANT: Weng, Zude  
 ; APPLICANT: Caimi, Perry G  
 ; APPLICANT: Anderson, Shawn  
 ; TITLE OF INVENTION: Plant Metabolism Genes  
 ; FILE REFERENCE: BB1378 US NA  
 ; CURRENT APPLICATION NUMBER: US/09/614,912  
 ; CURRENT FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: 60/143,401  
 ; PRIOR FILING DATE: 1999-07-12  
 ; PRIOR APPLICATION NUMBER: 60/143,412  
 ; PRIOR FILING DATE: 1999-07-12  
 ; PRIOR APPLICATION NUMBER: 60/146,650  
 ; PRIOR FILING DATE: 1999-07-30  
 ; PRIOR APPLICATION NUMBER: 60/170,906  
 ; PRIOR FILING DATE: 1999-12-15  
 ; PRIOR APPLICATION NUMBER: 60/172,959  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/172,946  
 ; PRIOR FILING DATE: 1999-12-21  
 ; NUMBER OF SEQ ID NOS: 204  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 144  
 ; LENGTH: 539  
 ; TYPE: PRT  
 ; ORGANISM: Triticum aestivum  
 ; FEATURE:  
 ; NAME/KEY: UNSURE  
 ; LOCATION: (272)..(273)

US-09-614-912-144

Query Match 10.1%; Score 334.5; DB 4; Length 539;  
Best Local Similarity 23.8%; Pred. No. 8.4e-27;  
Matches 120; Conservative 108; Mismatches 216; Indels 61; Gaps 15;

```
Qy      107 GTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNGPSFQK 166
          |  ||: |:|  ::| |  | |:|  ::| |:| ::| | :  :|
Db      2   GYIEGEITVSGYPKKQETFARISGYCEQNDIHSPhVTIYESLVFSAWLRRLPAEVDSERRK 61

Qy      167 K-VEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGLDC 225
          :| :|  : |: :  |:|  : |:| |:|::|| :|: :| ::  |||:|
Db      62 MFIEEIMDLVELTSLRGALVGLPGVNGLSTEQRKRLTIAVELVANPSIIFMDEPTSGLDA 121

Qy      226 MTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILS-FGELIFCG----TPAEM 280
          |  ::  :  | || ||||  ::|: ||:: ::  || |:|  | :
Db      122 RAAAIVMRTVRNTVNTGRTVVCTIHQPSIDIFEAFDELFLMKRGGEIYVGPVGQNSANL 181

Qy      281 LDFFNDG-----GYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMI-----ESA 328
          ::| :  ||  ||  :  :  |: :| : :|
Db      182 IEYFEEIEGISKIKDGY-----NPATWML-----EVSSSAQEEMLGIDFAEV 223

Qy      329 YKKAICHKTLKNIERMKHLKTLPM-----VPFKTKDSPGVFSLKGVLLRRVTRNLVRNK 383
          |::| :  :| | :| | ::|  : | |:|  ::  | :  : ||
Db      224 YRQSELYQ---RNKELIKEL-SMPAPGSSDLNFPTQYSRSFVTQCLACLWKQXXSYWRNP 279

Qy      384 LAVITRLLQNLIMGLFL--LFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNL 441
          ||| ::| |  :|: |  ::  ||  :  | | |:| : :
Db      280 SYTAVRLLFTIVIALMFGTMFWDLGSKTR----RSQDLFNAMGSMYAALYIGVQNSGSV 335

Qy      442 FPVL---RAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVAR 498
          ||:  | | :|  |:|  :  |: :| :|: :| :|  ||:
Db      336 QPVVVVVERTVFYRERAAGMYSAPFYAFGQVAIEFPYVLVQALIYGGVLVYSMIGFEWTVAK 395

Qy      499 FGYFSAALLAPHLIGEFLTLVLLGIVQNPNIVNSVVALLSIAGVLVG---SGFLRNIQEM 555
          | ::  :  |  |  :: :|: |  |: |:| |  |  ||:|  ::
Db      396 FLWYLFFMYFTMLYFTFYGMMAVGLTPN----ESIAAIISSAFYNVWNLFSGYLIPRPKL 451

Qy      556 PIPFKIISYFTFQKYCSEILVVNEF 580
          || ::  |:  :  || ::|
Db      452 PIWWRWYSWICPVAWTLYGLVASQF 476
```

RESULT 6

US-09-543-681A-5411

; Sequence 5411, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS  
MIRABILIS FOR

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 5411  
; LENGTH: 653  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-5411

Query Match 7.9%; Score 263.5; DB 4; Length 653;  
Best Local Similarity 22.0%; Pred. No. 5.5e-19;  
Matches 141; Conservative 111; Mismatches 203; Indels 187; Gaps 30;

```
Qy      68 ILKDVSLYVESGQIMCILGSSSGSKTLLDAMSGRLGRAGTFLGEVYVNGR----- 118
      :| :|| : :|::| :||:|||||:|::| : | | : : || | |:
Db      29 VLDQISLTINAGEMVAIIIGASGSGKSTLMNIL-GCLDKPSS--GEYKVAGQCQVADMESDQ 85

Qy     119 --ALRREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELS 176
      ||||| : :| | :||: || : : | : | : | :|| | | :
Db      86 LAALRREH----FGFIFQRYHLMHAHLTAEQNVEIPAIYAGK-----STEQRKERARALLT 136

Qy     177 LSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGLDCMTANQIVVLLV 236
      :|::| :| :| :||:||||| | : :|:| |||| || : :|| :|
Db     137 RLGLAERI--HYRPSQLSGGQQQRVSIARALMNGGEVILADEPTGALDSQSGKEVMAILK 194

Qy     237 ELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSN 296
      :| :| :||: | | : | | :| : :|| | | | | |
Db     195 QLNQQGHTVVIIVTHDPL--IAQQADRIEIKDGQII-----SDNNN-----HHS 237

Qy     297 PFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVFF 356
      | : : | : :| : :
Db     238 P-----VKKVPPAIQTASYFHQVI----- 256

Qy     357 KTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAI 416
      | |:: | :| ||: | :| :| :| :| :| :| :| :| :|
Db     257 -----GRFTQ---ALNMAWRAMVVNKIRTLTLM-LIIGIASVVTII-----VIGDAA 301

Qy     417 QDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQ-----ESQDGLYQKWQMMLAY 467
      :||| : :|| | :||:| || : | | | : :
Db     302 KDRVLADIKAI-----NTIDIYPGKELGSDSPEDKQSLTIQDQDALKQQSYIQ--- 351

Qy     468 ALHVLPPFSVVATMIFSS-----VCYWTGLGLHPEVARFGYFSAALL-----APHL 511
      || : ||| |:: : ||| | :
Db     352 -----SVTPQIYFSSRLRRGNQDAPATVSGVNED-----YFSVYALKFAQGSTFTPD 399

Qy     512 IGEFLTLVLLGIVQN-----PN---IVNSVVALLSIAGVLVG-----SGFLRNIQ-EM 555
      | :|| :| :| || : : :| :| :| | | :
Db     400 IHRQAQVVV--IDENTRHRFFPNQAVIGEQUIIRNIPSTIIGVVAEQKSTFGDNKSLRV 457

Qy     556 PIPFKIISYFTFQK-YCSEILV-VNEFYG-----LNFTCGSSNVSVTTNPMCAF 602
      :| :| : :| | | | | | | | : :
Db     458 WVPYSTLSSRIYNRSYLDNITVKVKEGYDASVAEQQILRLLTIRHGKKDI-----F 508

Qy     603 TQGIQFIEKTCPGATSRFTMNFLILYSFIPALVILGIVVFKI 644
      | | | | | : : :||: || | |
Db     509 TYNIDSFIAAEKTTQ--TMQLFLTLLVAVISLVGGIGVMNI 548
```

RESULT 7

US-09-489-039A-9127  
 ; Sequence 9127, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 9127  
 ; LENGTH: 384  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-9127

Query Match 7.8%; Score 259; DB 4; Length 384;  
 Best Local Similarity 23.9%; Pred. No. 6.7e-19;  
 Matches 84; Conservative 70; Mismatches 133; Indels 64; Gaps 11;

Qy	56	DITSCRQOWTR-QILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVY	114
		:  : :: :    :   :  :     : :        :  :   :	
Db	24	ELANIKKSFGRTQVLNDISLDIPSGQMVALLGPGSGKTTLLRIIAGLEHQTS---GHIR	80
Qy	115	VNGRALRREQFQD-CFSYVLQSDTLLSSLTVRETLHY--TALLAIRRGNPGSFQKKVEAV	171
		:  :   :  :    :   : : :       : :    :	
Db	81	FHGTDVSRMHARDRKVGFEVQHYALFRHMTVFDNIAFGLTVLPRRERPNAAAIKAKVTKL	140
Qy	172	MAELSLSHVADRLIGNYSLGGISTGERRRVISIAQLLQDPKVMLEFDEPTTGLDCMTANQI	231
		: :  : :    :   :: :    : ::       : :	
Db	141	LEMVQLAHLADRYPAQ-----LSGGQKQVALARALAVEPQILLDEPFGALDAQVRKEL	195
Qy	232	VVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAE-----MLD	282
		:  : : :   : :  :  :    :   :  :	
Db	196	RRWLRQLHEELKFTSVFVTHDQEEAMEVADRVVMSQGNIEQADAPERVWREPSTREVL	255
Qy	283	FFND-----CGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKR	321
		:      :     :       : :	
Db	256	FMGEVNRLQGVIKGGQFHVGAHRWPLGY-TPAYQGPVDFLRPWVVDI-SRRTSLDSPLP	313
Qy	322	VQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVL	372
		:: :      :       :  :	
Db	314	VQVLEASPK-----GHYTQLVVQPLGWYDEP-----LSVVL	344

RESULT 8

US-09-543-681A-8215  
 ; Sequence 8215, Application US/09543681A  
 ; Patent No. 6605709  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GARY BRETON  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS FOR

```

Query Match          7.8%; Score 258.5; DB 4; Length 210;
Best Local Similarity 33.7%; Pred. No. 2.6e-19;
Matches 70; Conservative 45; Mismatches 80; Indels 13; Gaps 4;

Qy          65 TRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQ 124
      |  || :|||::| | : | ||||| |||||:|::| | | | :
Db          11 TTGILTEVSLHLEQGCGCLGISGSSSGSGKTTLLNAIAGYTDYTGDI---VLANQNMMNKL PV 67

Qy          125 FQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHVADRL 184
      :| | | : | | | |||:: | | :: : ::| |:: : : ::| | |
Db          68 WQRPCRYLNQRLYLFPFLT VKQNWLWLAQYAAKQKRS----KEKEIALLEQMGIAHLATRY 123

Qy          185 IGNYSLGGISTGERRRVSIAAQLQDPKVM LFDEPTTGLDCMTANQIVVLLVELARRNRI 244
      || |||:| |::| | : ||::| ||| : || | | : |:: | ::
Db          124 PSQ-----ISGGEQQRVALARALISQPKLLLMDEPFSSLDWETRYQLWELIISLKKQOIT 178

Qy          245 VVLTIHQPRSELFQLFDKIAILSFGELI 272
      :: |::| || | || :|| |::
Db          179 MIIVTHEPR-ELQALADKTL LLSNGKIV 205

```

Query Match 7.6%; Score 253.5; DB 4; Length 373;  
Best Local Similarity 25.1%; Pred. No. 2.5e-18;

Matches 89; Conservative 61; Mismatches 133; Indels 71; Gaps 12;

```

Qy      44 SYSVSHRVRPWWDDITSCRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMS--- 100
      |  ::| |  ::| |  ::| ||:| | ||::| :|| ||||| |||  ::
Db      13 SIEINH-VTKYFDRT-----EVLHDVNLTVNSGEMMALLGPSGSGKTTLLRIIAGLE 63

Qy     101 ----GRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLLHY--TALL 154
      |::  ||  :  ::  |  :  :| |  |  :|| |  :  :  | |
Db      64 HQTEGKICFAGQDVSRLLHARERKV-----GFVFQHYALFRHMTVFENIAFGLTVLP 114

Qy     155 AIRRGNPGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVM 214
      | |  :  |||  ::  :  | ||| |  :| |::||::| |  :|::
Db     115 RRERPNKAAIDKKVTQLLEMIQLPHLAQRYPAQ-----LSGGQKQRVALARALAVEPQIL 169

Qy     215 LFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFC 274
      | |||  ||  ::  | ||  :  :  :  : |  :: | :| | :  ::
Db     170 LLDEPFGALDAKVRTELRSWLRELHSELKFTSVFVTHDQQEAMEVADRIVIMGNKGIEQV 229

Qy     275 GTPAE-----MLDFFND-----CGYPCPEHSNPFDFYMDLTSVDTQ 310
      ||| :  :|:| |  || |  |  |  |  ||
Db     230 GTPQQVWHTPESRFVLEFLGDVNHQLQGEINGAQLQIAGYHPLSVTP--LYQG--KVDVF 285

Qy     311 SKEREIETSKRVQMIESAYKKSACHKTLKNIE--RMKHLKTLPMVPFKTKDSP 362
      :  ||  :  :|  :| ||  |  |  :  |  :  ::|
Db     286 LRPWEISLNPH-----SDSLCKLPVKVIEVTPKGYHWQLVLQPIEWGNTP 330

```

RESULT 10

US-09-540-236-3618

; Sequence 3618, Application US/09540236

; Patent No. 6673910

; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540,236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 3618

; LENGTH: 245

; TYPE: PRT

; ORGANISM: M.catarrhalis

US-09-540-236-3618

Query Match 7.5%; Score 249.5; DB 4; Length 245;

Best Local Similarity 32.3%; Pred. No. 3.2e-18;

Matches 74; Conservative 49; Mismatches 87; Indels 19; Gaps 9;

```

Qy      62 QQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSG--RLGRAGTFLGEVYVNGRA 119
      |:|  :::| |  :| ||:: ||| :|:| | |  :  :  ||:  ::  |
Db      19 QRW---VVEDVSFEIEQGQVVGILGPNAGKTTSFYMVIGLVPMDKGQVILGDQDISKNA 75

Qy     120 LRREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQK-KVEAVMAELSL 178
      :  |:  :|  :|  ::  |||  :  :  |:|  |:  :  |:  ::|  :|:  |
Db     76 M-HERAAKGIGYLPQEASIFRKLTV EQNI--MAILQTRKDLTQTEQRQQLKLMADFHLE 132

```



Qy 179 HVADRLIGNYSLG-GISTGERRRVSIQAQLLQDPKVMFLFDEPTTGLDCMTANQIVVLIVE 237  
 || :||| :| |||| || | :|| :| || | :| :: : | : ::  
 Db 133 HV-----RHSLGMSVSGGERRRCEIARCLASNPKEILLDEPFAGVDPISVSDIMQVIET 186

Qy 238 LARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND 286  
 | | | :| | | | : | | :| | :| | | :| || :  
 Db 187 LRERGIGVLITDHNVR-ETLSICQKAYIVSEGKVIAQGNKDEIL--FNE 232

RESULT 11

US-09-489-039A-13987  
 ; Sequence 13987, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 13987  
 ; LENGTH: 344  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-13987

Query Match 7.4%; Score 245; DB 4; Length 344;  
 Best Local Similarity 31.7%; Pred. No. 1.8e-17;  
 Matches 84; Conservative 44; Mismatches 89; Indels 48; Gaps 12;

Qy 42 HASYSVSHRVRPWWDDITSCRQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSG 101  
 | | | | : | | | | :| :| | | | :| | :| :| | :  
 Db 13 HVSKSFSRKGHP-----VLALQHINLSIERGDIFGIIGYSGAGKSTLL-RLIN 59

Qy 102 RLGRAGTFLGEVYVNGRALRREQFQDC-----FSYVLQSDTLLSSLTVRETLHY 150  
 || | ||| :|| | ||| : | : || :| || |  
 Db 60 RLETPGE--GEVLLNG-----EPLQDCSGQRLQAIKKDIGMIFQNFNLLNSKTV---FHN 109

Qy 151 TALLAIRRGNPGSF-QKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIQAQLLQ 209  
 | : | :| :| | :| :| :| | :| :| :| | :| :| :| |  
 Db 110 IAIPILQGRDKAFIQARVAELLAFAVDLS---DK-IHSYP-NELSGGQKQRVGIARALAT 164

Qy 210 DPKVMFLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFG 269  
 :| | :| || | :| | | || :| | | :| :| | :| :| :| |  
 Db 165 NPSVLLCDEATSALDPHTTVQILLLLQEINRRYGITIVLITHEMSVIQKICHKVAVMQAG 224

Qy 270 ELIFCGTPAEMLDFFNDCGYPCPEH 294  
 : : | : | | | :|  
 Db 225 RIVEQGA---VFDLFAQ-----PQH 241

RESULT 12



```

; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,447
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5994088nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 95,1121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-9808
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-752-447-2

```

```

Query Match          7.3%; Score 244; DB 2; Length 1280;
Best Local Similarity 20.7%; Pred. No. 2.3e-16;
Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

```

```

Qy      41 LHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMS 100
      :| ||      |:      ||| ::| |:| | : ::|:| | |:| : |
Db      397 VHFSYPSRKEVK-----ILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQ 441

Qy      101 GRLGRAGTFLGEVYVNGRALR--REQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRR 158
      ||      | | |:| :|      :|      |:      : :| : | : |
Db      442 -RL--YDPTEGMVSVGDQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRY----- 490

Qy      159 GNPGSFQKKVEAVMAE-----LSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDP 211
      |      ::| : |      : | | | |:|      :| |::|:|:| |::|
Db      491 GRENVTMDEIEKAVKEANAYDFIMKLPKFDTLVGERG-AQLSGGQKQRIAIARALVRNP 549

Qy      212 KVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIIHQPRSELFQLFDKIAILSFGE 271
      |:| | | |:| | : :| : ::| |:| : : | | : | | | : |
Db      550 KILLDEATSALD-TESEAVVQVALDKARKGRTTIVIAH--RFATVRNADVIAGFDDGVI 606

Qy      272 IFCGTPAEMLDFFNDGYPCEHSNPFDFYMDLTSVDTQSKEREIETSK----- 320
      : | |:| : :      | | ::| | | |:| :
Db      607 VEKGNHDELM-----KEKGIYFKLVTMQTAGNEVELENAADESKSEIDALE 652

Qy      321 -----RVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPF-----KTKDSPG 363
      | :      : : | |:| | : : ::| | | : |
Db      653 MSSNDSRSSLIKRSTRRSVRGSQAQHRKLSTKEALD--ESIPPVSFWRIMKLNLTWPY 710

Qy      364 -----VFSK-LGVLLR-----RVTRN-----LVRNKLAVITR 389
      :||| :| | |      | |      | :| |

```

Db 711 FVVGVFCAIINGGLQPAFAIIFS KIIGVFTRIDDPETKRQNSNLFSLFLALGIISFITF 770

Qy 390 LLQNLMG-----LFLFFVLRVRSNV-----LKGAIQ 417  
 || | : | : | : |||

Db 771 FLQGFTFGKAGEILTKRLRYMVFRSMLRQDVSWFHDPKNTTGALTTRLANDAAQVKGAIG 830

Qy 418 DRV-----GLLYQFVGATPYTGMLNAVNLFVPL-----RAVSDQE 452  
 |: |: |: |: |: |: |: |: |: |

Db 831 SRLAVITQNIANLGTGIIISFIYGWQLTLLLLAI--VPIIAIAGVVMKMFAGQALKDKK 888

Qy 453 SQDGL-----YQKWQMMLAYALHV-----LPFSVVATM 480  
 :| ||: ||: | :| | : || |

Db 889 ELEGAGKIAATEAIENFRTVVSLTQEQQFEHMYAQSLOQVPYRNSLRKAHIFGITFSFTQAM 948

Qy 481 IFSSV--CYWTLGLHPEVARFGYFSAALLAPHLIGEF--LTLVLLGIVQNPNI VNSV--- 533  
 :: | |: || | | : | : | : | |

Db 949 MYFSYAGCF-----RFG---AYLVAHKLMSFEDVLLVFSAVVFGAMAVGQVSSF 994

Qy 534 -----VALLSIAGVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVNEFYGLNFTCG 588  
 | :| | :: |:: |: | :| | : | | | |

Db 995 APDYAKAKISAAHIIM-----IIKTPL-----IDSYSTEGLMPNTLEG-NVTFG 1038

Qy 589 SSNVSVTTNPMCAFTQGI 606  
 : | | ||:

Db 1039 EVVFNYPTRPDIPVLQGL 1056

RESULT 14

US-09-316-167-2

; Sequence 2, Application US/09316167

; Patent No. 6365357

; GENERAL INFORMATION:

; APPLICANT: Mechetner, Eugene

; APPLICANT: Roninson, Igor B

; TITLE OF INVENTION: Methods and Reagents for Preparing and

; TITLE OF INVENTION: Using Immunoligcal Agents Specific for P-glycoprotein

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.

; STREET: 300 South Wacker Drive, Seventh Floor

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/316,167

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/752,447

; FILING DATE: 15-NOV-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6365357nan, Kevin E  
 ; REGISTRATION NUMBER: 35,303  
 ; REFERENCE/DOCKET NUMBER: 95,1121  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-913-0001  
 ; TELEFAX: 312-913-9808  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1280 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-09-316-167-2

Query Match 7.3%; Score 244; DB 4; Length 1280;  
 Best Local Similarity 20.7%; Pred. No. 2.3e-16;  
 Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

Qy	41	LHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMS	100
		:      :	
Db	397	VHFSYPSRKEVK-----ILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQ	441
Qy	101	GRLGRAGTFLGEVYVNGRALR--REQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRR	158
		:   :   :   :   :	
Db	442	-RL--YDPTEGMVSVGDQDIRTINVRLREIIGVVSQEPVLFATTIAENIRY-----	490
Qy	159	GNPGSFQKKVEAVMAE-----LSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDP	211
		:   :   :       :   :   :   :	
Db	491	GRENVTMDEIEKAVKEANAYDFIMKLPHKFDTLVGERG-AQLSGGQKQRIAIARALVRNP	549
Qy	212	KVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGEI	271
		:        :    : :   :    :   :   :   :   :	
Db	550	KILLLLDEATSALD-TESEAVVQVALDKARKGRTTIVIAH--RFATVRNADVIAGFDDGVI	606
Qy	272	IFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSK-----	320
		:     :   :   :   :   :	
Db	607	VEKGNHDELM-----KEKGIYFKLVMTQTAGNEVELENAADESKSEIDALE	652
Qy	321	-----RVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPF-----KTKDSPG	363
		: : :     :   : :   :   :	
Db	653	MSSNDSRSSLIRKRSTRRSVRGSQAQHRKLSTKEALD--ESIPPVSFWRIMKLNLEWPHY	710
Qy	364	-----VFSK-LGVLLR-----RVTRN-----LVRNKLAVITR	389
		:      :          :   :	
Db	711	FVVGVFCAIINGGLQPAFAIIFSKIIGVFTRIDDPETKRQNSNLFSLFLALGIISFITF	770
Qy	390	LLQNLMG-----LFLFFVLRVRSNV-----LKGAIQ	417
		:   :   :   :   :	
Db	771	FLQGFTFGKAGEILTKRLRYMVFRSMLRQDVSWFHDPKNTTGALTTRLANDAAQVKGAI	830
Qy	418	DRV-----GLLYQFVGATPYTGMLNAVNLFPVL-----RAVSDQE	452
		:   :   :   :   :   :   :	
Db	831	SRLAVITQNIANLGTGIIISFIYGWQLTLLLLAI--VPPIAIAAGVVEMKMFAGQALKDKK	888
Qy	453	SQDGL-----YQKWQMMLAYALHV-----LPFSVATM	480
		:     :     :     :	
Db	889	ELEGAGKIAIEAIENFRTVVSLTQEQKFEHMYAQSLQVPYRNSLRKAHIFGITFSFTQAM	948

Qy 481 IFSSV--CYWTLGLHPEVARFGYFSAALLAPHLIGEF--LTLVLLGIVQNPNIIVNSV--- 533  
 :: | | : || | | : | : || : | |  
 Db 949 MYFSYAGCF-----RFG----AYLVAHKLMSFEDVLLVFSASVVGAMAVGQVSSF 994  
 Qy 534 -----VALLSIAGVLVGSGFLRNIQEMPPIPKIISYFTFQKYCSEILVVNEFYGLNFTCG 588  
 | : | | :: | : | : | : | | | |  
 Db 995 APDYAKAKISAAHIIM-----IIEKTPL-----IDSYSTEGLMPNTLEG-NVTFG 1038  
 Qy 589 SSNVSVTTNPMCAFTQGI 606  
 : | | ||:  
 Db 1039 EVFENYPTRPDIPVLQGL 1056

RESULT 15

US-09-397-233-2

; Sequence 2, Application US/09397233

; Patent No. 6630327

; GENERAL INFORMATION:

; APPLICANT: Mechetner, Eugene

; Roninson, Igor B

; TITLE OF INVENTION: Methods and Reagents for Preparing and  
 ; Using Immunological Agents Specific for P-  
 glycoprotein

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff

; STREET: 300 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/397,233

; FILING DATE: 16-Sep-1999

; CLASSIFICATION: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 6630327nan, Kevin E

; REGISTRATION NUMBER: 35,303

; REFERENCE/DOCKET NUMBER: 95,1121-C

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312-913-0001

; TELEFAX: 312-913-0002

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1280 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-397-233-2

Query Match 7.3%; Score 244; DB 4; Length 1280;  
Best Local Similarity 20.7%; Pred. No. 2.3e-16;  
Matches 153; Conservative 106; Mismatches 229; Indels 250; Gaps 32;

```
Qy      41 LHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSSGKTTLLDAMS 100
      :| ||      |:      ||| ::| |:||| : ::||| ||:| : |
Db      397 VHFSYPSRKEVK-----ILKGLNLKVQSGQTVALVGNSGCGKSTTVQLMQ 441

Qy      101 GRLGRAGTFLGEVYVNGRALR--REQFQDCFSYVLQSDTLLSSLTVRETLYHTALLAIRR 158
      ||      | | |:|: :| :|      |: : :| : | : |
Db      442 -RL--YDPTEGMVSVGDQDIRTINVRFLREIIGVVSQEPVLFATTIAENIRY----- 490

Qy      159 GNPGSFQKKVEAVMAE-----LSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDP 211
      |      ::| : |      : | | | |:|      :| |::|:|:| |::|
Db      491 GRENVTMDEIEKAVKEANAYDFIMKLPKFKDTLVGERG-AQLSGGQKQRIAIARALVRNP 549

Qy      212 KVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGEL 271
      |:| || |: || : :| : :: ||: | :: | | : | || | :
Db      550 KILLLDEATSALD-TESEAVVQVALDKARKGRTTIVIAH--RFATVRNADVIAGFDDGVI 606

Qy      272 IFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSK----- 320
      : |      |:|      | | :: | | |:| :
Db      607 VEKGNHDELM-----KEKGIYFKLVMTQTAGNEVELENAADESKSEIDALE 652

Qy      321 -----RVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPF-----KTKDSPG 363
      | :      : : | | : | | :      ::| | |      : |
Db      653 MSSNDSRSSLIRKRSTRRSVRGSQAQHRKLTKEALD--ESIPPVSWFRIMKLNLTWPY 710

Qy      364 -----VFSK-LGVLLR-----RVTRN-----LVRNKLAVITR 389
      :||| :|| |      | |      |      :: ||
Db      711 FVVGVFCAIINGGLQPAFAIIFSKIIGVFTRIDDPETKRQNSNLFSLFLALGIISFITF 770

Qy      390 LLQNLMG-----LFLFFVLRVRSNV-----LKGAIQ 417
      || |      | : | :| :|      :|||
Db      771 FLQGFTFGKAGEILTKRLRYMVFRSMLRQDVSWFHDPKNTTGALTTRLANDAAQVKGAIG 830

Qy      418 DRV-----GLLYQFVGATPYTGMLNAVNLFPVL-----RAVSDQE 452
      |:      |:| |: | :| |: |::      :|: |:|
Db      831 SRLAVITQNIANLGTGIIISFIYGWQLTLLLLAI--VPIIAIAGVVEMKMFAGQALKDKK 888

Qy      453 SQDGL-----YQKWQMMLAYALHV-----LPFSVVATM 480
      :|      ||:: | | :| |      : || |
Db      889 ELEGAGKIAIEAIENFRTVVSILTQEQQFEHMYAQSLQVPYRNSLRKAHIFGITFSFTQAM 948

Qy      481 IFSSV--CYWTLGLHPEVARFGYFSAALLAPHLIGEF--LTLVLLGIVQNPINIVNSV--- 533
      :: | |:      ||| | | : | : || :| | |
Db      949 MYFSYAGCF-----RFG----AYLVAHKLMSFEDVLLVFSAVVFGAMAVGQVSSF 994

Qy      534 -----VALLSIAGVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCG 588
      | :| | ::|      |:| :|      | :| |:| | | | |
Db      995 APDYAKAKISAAHIIM-----IIEKTPL-----IDSYSTEGLMPTLEG-NVTFG 1038

Qy      589 SSNVSVTTNPMCAFTQGI 606
      : | |      ||:
Db      1039 EVVFNYPTRPDIPVLQGL 1056
Search completed: February 27, 2004, 07:20:16
Job time : 16.7508 secs
```

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:44:33 ; Search time 14.5049 Seconds  
(without alignments)  
4317.206 Million cell updates/sec

Title: US-09-989-981A-6  
Perfect score: 3326  
Sequence: 1 MGDLSLTPGGSMGLQVNRG.....PALVILGIWVFKIRDHLISR 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	668.5	20.1	656	2	JC7860	brain multidrug re
2	649.5	19.5	725	2	T47652	ABC transporter-li
3	627	18.9	1294	2	S77690	probable membrane
4	621	18.7	1049	1	S19421	ATP-dependent perm
5	614	18.5	720	2	T47648	ABC transporter-li
6	610.5	18.4	725	2	C84423	probable ABC trans
7	608	18.3	708	2	T47650	ABC transporter-li
8	602.5	18.1	687	1	FYFFW	white protein - fr
9	597	17.9	590	2	B96573	protein F12M16.17
10	595.5	17.9	658	2	T31958	hypothetical prote
11	591.5	17.8	740	1	T02567	probable ATP-bindi
12	590.5	17.8	638	2	G02068	white homolog - hu
13	589.5	17.7	646	2	C86441	probable ABC trans



14	584	17.6	755	2	G84791	probable ABC trans
15	583.5	17.5	687	2	D96553	hypothetical prote
16	578.5	17.4	646	2	JC7777	ATP binding casset
17	575.5	17.3	649	2	A84509	probable ABC trans
18	572	17.2	739	2	T45891	ABC transporter-li
19	571.5	17.2	609	2	E96742	probable ABC trans
20	564	17.0	662	2	T47649	ABC transporter-li
21	554	16.7	559	2	B88474	protein C05D10.3 [
22	552.5	16.6	635	2	T08934	hypothetical prote
23	545.5	16.4	608	2	T34391	hypothetical prote
24	540	16.2	678	2	H96552	hypothetical prote
25	539	16.2	577	2	T04229	ABC-type transport
26	529.5	15.9	633	2	T19189	hypothetical prote
27	527.5	15.9	659	2	E86313	hypothetical prote
28	517.5	15.6	610	2	T19333	hypothetical prote
29	517.5	15.6	639	2	G88839	protein C10C6.5 [i
30	514.5	15.5	705	2	D84680	probable ABC trans
31	504	15.2	695	2	T21109	hypothetical prote
32	497	14.9	547	2	T31543	hypothetical prote
33	488	14.7	737	2	T46101	ABC transporter-li
34	464	14.0	1450	2	A84780	probable ABC trans
35	456.5	13.7	1426	2	T30567	ATP-binding casset
36	454.5	13.7	675	1	FYFFB	brown protein - fr
37	435	13.1	668	2	S55023	brown protein - fr
38	429	12.9	1469	2	H96622	probable ABC trans
39	425.5	12.8	1443	2	T02491	probable ABC trans
40	425	12.8	1451	2	B86286	F9L1.15 protein -
41	421	12.7	1450	2	T45888	ABC transporter-li
42	420	12.6	1435	2	D96693	protein Putative A
43	418.5	12.6	1420	2	T02644	ABC-type transport
44	417.5	12.6	1619	2	T30541	ABCl transport pro
45	413	12.4	1413	2	G84790	probable ABC trans

# ALIGNMENTS

## RESULT 1

JC7860

brain multidrug resistance protein, BMDP - pig

C;Species: Sus scrofa domestica (domestic pig)

C;Date: 18-Nov-2002 #sequence\_revision 18-Nov-2002 #text\_change 31-Mar-2003

C;Accession: JC7860

R;Eisenblaetter, T.; Galla, H.J.

Biochem. Biophys. Res. Commun. 293, 1273-1278, 2002

A;Title: A new multidrug resistance protein at the blood-brain barrier.

A;Reference number: JC7860; MUID:22050127; PMID:12054514

A;Accession: JC7860

A;Molecule type: mRNA

A;Residues: 1-656 <EIS>

A;Cross-references: GB:AJ420927

A;Experimental source: brain

C;Comment: This protein, a new transport protein of the ATP-binding cassette (ABC) superfamily of transporters, expressed in porcine brain capillary endothelial cells, plays an importnat role in the exclusion of xenobiotics from the brain and participates in drug transport across the blood-brain barrier and therefore is considered as a efflux pump at the cerebral endothelium.

C;Genetics:  
A;Gene: bmdp

Query Match 20.1%; Score 668.5; DB 2; Length 656;  
Best Local Similarity 28.5%; Pred. No. 3.4e-42;  
Matches 180; Conservative 144; Mismatches 252; Indels 55; Gaps 18;

```
Qy      13 MGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDV 72
      : : : : : | | : : : | : : : : : | | : : | :
Db      8 VSIPMSKRNTNGLPGSSSNELKTSAGGAVLSFHDICYRVKVKSGFLFCRKTVEKEILTNI 67

Qy      73 SLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYV 132
      : : : | : | | : | | : | | : : | | : | | : | |
Db      68 NGIMKPG-LNAILGPTGGGKSSLLDVLAAKDPHG-LSGDVLING-APRPANFKCNSGYV 124

Qy     133 LQSDTLLSSLTVRETLHYTALLAIRRGNPGSF-----QKKVEAVMAELSLSHVADRLIGN 187
      : | | : : | | | | | : : | | : : : : | | | | | : |
Db     125 VQDDVVMGTLTVRENLOFSAALRL----PTMTNHEKNERINMVIQELGLDKVADSKVGT 180

Qy     188 YSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVL 247
      : | : | | | : | | : | | : : | | | | | | | | | : : : | :
Db     181 QFIRGVSGGERKRTSIAMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMSKQGRTIIF 240

Qy     248 TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTSV 307
      : | | | | : | : | | : : : | | : | | | : | | : | : :
Db     241 SIHQPRYSIFKLFDSLTLASGRMLFHGPAREALGYFASIGYNCEPYNNPADFFLDVING 300

Qy     308 DTQ-----SKEREIETSKRVQMIE---SAYKKSACHKTLKNIE-----RMK 346
      | : : : : : : | | : | : : : | : : : : : : :
Db     301 DSSAVVLSRADRDEGAQEPEEPPEKDTPLIDKLAIFYTNSSFFKDTKVELDQFSGGRKKK 360

Qy     347 HLKTLPMVPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMGLFL--LFFV 404
      | : | : | : : | : | : | : | : | : : : : | : : :
Db     361 KSSVYKEVTTYTT---SFCHQLRWISRRSFKNLLGNPQASVAQIIVTIILGLVIGAIIFYD 416

Qy     405 LRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMM 464
      | : | | | : | : | : : : | | : : : | | : :
Db     417 LK---NDPSG-IQNRAGVLF-FLT TNQCFSSVSAVELLVVEKKLFIHEYISGYRVSSYF 471

Qy     465 LAYAL-HVLPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGI 523
      | : | : : : | : : | | | | | : : : : : | |
Db     472 FGKLLSDLLPMRMLPSIIFTCITYFLLGLKPAVGSGFFIMMFTLM---MVAYSASSMALAI 528

Qy     524 VQNPNIIVNSVVALLSIAGV--LVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFY 581
      : : : | : : : | : : | | : : : : | : : | | |
Db     529 AAGQSVVSVATLLMTISFVFMIFSGLLVNLKTVPWLSWLQYFSIPRYGFSALQYNEFL 588

Qy     582 GLNFTCGSSNVSVTTNPMCAFT--QGIQFIE 610
      | | | : : | | : | : : | : :
Db     589 GQNFCPG---LNVTTNNTCSFAICTGAEYLE 616
```

RESULT 2

T47652

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T26I12.10

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000  
 C;Accession: T47652  
 R;Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.;  
 Mayer, K.F.X.; Quetier, F.; Salanoubat, M.  
 submitted to the Protein Sequence Database, February 2000  
 A;Reference number: Z24471  
 A;Accession: T47652  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-725 <MON>  
 A;Cross-references: EMBL:AL132954  
 A;Experimental source: cultivar Columbia; BAC clone T26I12  
 C;Genetics:  
 A;Map position: 3  
 A;Note: T26I12.10  
 C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein  
 Fl2L6.1; ATP-binding cassette homology

Query Match 19.5%; Score 649.5; DB 2; Length 725;  
 Best Local Similarity 29.4%; Pred. No. 1.1e-40;  
 Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;

Qy	33	PEPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSGSGK	92
		:   :     :       : :           :   :   :   :	
Db	68	PVPYVLNFNNLQYDVTLR--FGFSRQNGVKTLLDDVSGEASGDILAVLGASGAGK	123
Qy	93	TTLLDAMSGRLGRAGTFLGEVYVNG-RALRREQFQDCFSYVLQSDTLLSSLTVRETLHYT	151
		:   :   :   :   :   :   :   :   :   :   :   :   :   :	
Db	124	STLIDALAGRAE-GSLRGSVTLNGEKLQSRLLKVISAYVMQDDLLFPM LTVKETLMFA	182
Qy	152	ALLAIRRG-NPGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLQD	210
		: :   : :   :   :   :   :   :   :   :   :   :   :	
Db	183	SEFRLPRSLSKSKMERVEALIDQLGLRNAANTVIGDEGHRGVSGGERRRVSIGIDIHD	242
Qy	211	PKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGE	270
		:       :       :   :   :   :   :   :   :   :   :	
Db	243	PIVFLFDEPTSGLDSTNAFMVVQVLKRIAQSGSIVIMSIHQPSARIVELLDRLIILSRGK	302
Qy	271	LIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETS-----	319
		:     :   :   :   :   :   :   :   :   :   :   :	
Db	303	SVFNGSPASLPGFFSDFGRPIPEKENISEFALDLV-----RELEGSNEGKALVDEN	354
Qy	320	-----KRVQMIESAYK-----KSAICKTL--KNIERMKHLKTLPMVPFKTKD	360
		: : :   :   :   :   :   :   :   :   :   :   :	
Db	355	EKWQONKISLIQSAPQTNKLDQDRSLSLKEAINASVSRGKLVSGSSRSNPTSMETVSSYA	414
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMGLFLLFFVLRVRSNVLKGAIQDRV	420
		:   :   :   :   :   :   :   :   :   :   :   :   :	
Db	415	NPSLFETF-ILAKRYMKNWIRMPVLVGTRIATVMVTGC-LLATVYWKLDHTPRGA-QERL	471
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLPPFSVATM	480
		:     :   :   :   :   :   :   :   :   :   :   :	
Db	472	-TLFAFVVPTMFYCCLDNVPVFIQERYIFLRETTNAYRTSSYVISHSLVSLPQLLAPSL	530
Qy	481	IFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIVNS-VVALLSI	539
		:   : : :   :   :   :   :   :   :   :   :   :   :	
Db	531	VFSAITFWTVGLSGGLEGFVYCLLIYASFWSGSSVVTFISGVV--PNIMLCYMVSTITYL	588

```

Qy      540 AGVLVSGGFLRNIQEMPIPFKIIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599
        |  |: ||| |  :| :  | : || | ::|||  :|
Db      589 AYCLLLSGFYVNRDRIPFYWTFWHYISILKYPYEAVLINEF-----DDPS 633

Qy      600 CAFTQGIQFIEKTCPGATS 618
        | :|:| : | | |
Db      634 RCFVRGVQVFDSTLLGGVS 652

```

# RESULT 3

S77690

probable membrane protein YOL075c - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein O1125; hypothetical protein O1130; hypothetical protein YOL074c

C;Species: *Saccharomyces cerevisiae*

C;Date: 21-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 19-Apr-2002

C;Accession: S77690; S66767; S66768

R;Alexandraki, D.; Katsoulou, C.; Tzermia, M.

submitted to the Protein Sequence Database, July 1996

A;Reference number: S66756

A;Accession: S77690

A;Molecule type: DNA

A;Residues: 1-1294 <ALE>

A;Cross-references: EMBL:Z74816; MIPS:YOL075c

A;Note: this is a revision to the sequence from reference S66756

A;Accession: S66767

A;Molecule type: DNA

A;Residues: 1-179, 'TTRTGVFLVVKRED' <ALW>

A;Cross-references: EMBL:Z74816

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be protein YOL074c

A;Accession: S66768

A;Molecule type: DNA

A;Residues: 200-1294 <ALF>

A;Cross-references: EMBL:Z74817

A;Experimental source: strain S288C

A;Note: this sequence has been revised in reference S77690

A;Note: this was assumed to be the complete sequence of protein YOL075c

C;Genetics:

A;Cross-references: SGD:S0005435

A;Map position: 15L

A;Note: YOL075c

C;Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop; transmembrane protein

F;45-263/Domain: ATP-binding cassette homology <ABC1>

F;62-69/Region: nucleotide-binding motif A (P-loop)

F;376-392/Domain: transmembrane #status predicted <TM1>

F;469-485/Domain: transmembrane #status predicted <TM2>

F;496-512/Domain: transmembrane #status predicted <TM3>

F;606-622/Domain: transmembrane #status predicted <TM4>

F;710-916/Domain: ATP-binding cassette homology <ABC2>

F;727-734/Region: nucleotide-binding motif A (P-loop)

F;1042-1058/Domain: transmembrane #status predicted <TM5>

F;1125-1141/Domain: transmembrane #status predicted <TM6>

F;1177-1193/Domain: transmembrane #status predicted <TM7>  
F;1269-1285/Domain: transmembrane #status predicted <TM8>

Query Match 18.9%; Score 627; DB 2; Length 1294;  
Best Local Similarity 31.7%; Pred. No. 1.2e-38;  
Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

```
Qy      65 TRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFL-----GEVYVNGRA 119
      |::||: |:  : | | |:| ||||:|::: :||| :|  |  | : |
Db      706 TKEILQSVNAIFKPGMINAIMGPSGSGKSSLLNLISGRL-KSSVFAKFDTSGSIMFNDIQ 764

Qy      120 LRREQFQDCFSYVLQSDT-LLSSLTVRETlHYTALLAIRRGNPGSFQKKVEAVMAELSLS 178
      :  |:: ||| | | ||::|||:||| | | | :  : : : : | |
Db      765 VSELMFKNVCSYVSQDDDHLLAALTvkETLKYAAALRLHHLTEAERMERTDNLIRSLGLK 824

Qy      179 HVADRLIGNYSLGGISTGERRRVsIAAQLLDPKVMLFDEPTTGLDCMTANQIVVLLVEL 238
      | : :||| : ||| ||:||||: ||| || :| |||:||| |: |: :| :|
Db      825 HCENNIIGNEFVKGISGGEKRRVTMGVQLLNDPPILLLDEPTSGLDsFTSATILEILEKL 884

Qy      239 AR-RNRIVVLTlHQPRSELFQlFDKIAILS-FGELIFCGTPAEMLDFFNDCGYPCPEHSN 296
      | : : : :||| |||||:| : :| : | |:| ||: :| : || || :|
Db      885 CREQGKTIIITlHQPRSELFKRFGNVLLLAKSGRTAFNGSPDEMIAYFTELGYNCPSFTN 944

Qy      297 PFDfYMDLTsVDTQSKEREIETSKRVQMIESAYKKSAICHKTLKNIERMKHLKTLPMVPF 356
      ||::|| ||:||::: || : ||: | ||:| |:: : | |:
Db      945 VADFFLDLISVNTQNEQNEISSRARVEKILSAWK-----ANMDN-ESLSPTPISEK 994

Qy      357 KTKDSPGVFSKLGVLlRRVTRNLV-----RNKLAVITRLlQNlIMGLFLL 401
      :  |:: :|: ||| |::: |: | :|:
Db      995 QQYSQESFFTEYSEFVRK-PANLVLAYIVNVKRQFTTTRRSFDsLMARIAQIPGLGVIFA 1053

Qy      402 FFVLRVRsNVlKGAIQDRVGLLYQfVGATPYTGMLNAVNLFPVLRAVSDQESQDGLYQKW 461
      | |: | :|:|:| : | : ||| : :| | :| | :|
Db      1054 LFFAPVKHNYT--SISNRLGLAQEST-ALYfVGMlGNLACYPTERDYfYEEYNDNVYGIA 1110

Qy      462 QMMLAYALHVLpFSVVATMIFSSVCYWTlGLHPEVARFGYfSAALLAPHLI---GEFLTL 518
      ||| || | :|:::| || | | | | : : || | :
Db      1111 PFFLAYMTLElPLSALASVLYAVFTVLACGL-PRTA--GNFFATVYCSFIVTCCGERLGI 1167

Qy      519 VLLGIVQNPN-IVNSVVALLSIAGVLVGSgFLRNIQEMPIPFKIISYFTfQKYCSEILVW 577
      :  : | :|| : :||| : | | | :| | | | :| | | :|
Db      1168 MTNTFFERPGfVvNCISIIlSIGTQMSGLMSL----GMSRVlKGfNYlNPVGYTSMIIN 1223

Qy      578 NEfYG-LNFTC--GSSNVSVTTNPMCAFTQG 605
      | | | || | | | | | |
Db      1224 FAfPGNLKlTCEDGGKNSDGT---CEfANG 1250
```

RESULT 4  
S19421

ATP-dependent permease ADP1 precursor - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: protein YCR011c; protein YCR105

C;Species: *Saccharomyces cerevisiae*

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 19-Jan-2001

C;Accession: S19421; S40914

R;Goffeau, A.; Purnelle, B.; Skala, J.

submitted to the Protein Sequence Database, March 1992

A;Reference number: S19420  
A;Accession: S19421  
A;Molecule type: DNA  
A;Residues: 1-1049 <GOF>  
A;Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42328.1; PID:g1907154;  
GSPDB:GN00003; MIPS:YCR011c  
R;Purnelle, B.; Skala, J.; Goffeau, A.  
Yeast 7, 867-872, 1991  
A;Title: The product of the YCR105 gene located on the chromosome III from  
Saccharomyces cerevisiae presents homologies to ATP-dependent permeases.  
A;Reference number: S40914; MUID:92160395; PMID:1789009  
A;Accession: S40914  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-1049 <PUR>  
R;Skala, J.; Purnelle, B.; Goffeau, A.  
Yeast 8, 409-417, 1992  
A;Title: The complete sequence of a 10.8 kb segment distal of SUF2 on the right  
arm of chromosome III from Saccharomyces cerevisiae reveals seven open reading  
frames including the RVS161, ADP1 and PGK genes.  
A;Reference number: S25353; MUID:92327849; PMID:1626432  
A;Contents: annotation  
C;Genetics:  
A;Gene: SGD:ADP1; MIPS:YCR011c  
A;Cross-references: SGD:S0000604; MIPS:YCR011c  
A;Map position: 3R  
C;Superfamily: ATP-dependent permease ADP1; ATP-binding cassette homology  
C;Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein  
F;1-25/Domain: signal sequence #status predicted <SIG>  
F;26-1049/Product: ATP-dependent permease ADP1 #status predicted <MAT>  
F;26-324/Domain: extracellular #status predicted <EXT>  
F;325-341/Domain: transmembrane #status predicted <TM1>  
F;406-607/Domain: ATP-binding cassette homology <ABC>  
F;423-430/Region: nucleotide-binding motif A (P-loop)  
F;550-557/Region: nucleotide-binding motif B  
F;794-810/Domain: transmembrane #status predicted <TM2>  
F;829-845/Domain: transmembrane #status predicted <TM3>  
F;878-894/Domain: transmembrane #status predicted <TM4>  
F;909-925/Domain: transmembrane #status predicted <TM5>  
F;938-954/Domain: transmembrane #status predicted <TM6>  
F;1025-1041/Domain: transmembrane #status predicted <TM7>  
F;50,114,165,221/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;429/Binding site: ATP (Lys) #status predicted

```

      |||      ||| ||:||||| :||: | |: |||:|||  || :: || |:  ||
Db      522 IGNEFDRGISGGEKRRVSIACELVTSPLVFLDEPTSGLDASNANNVIECLVRLSSDYNR 581
Qy      244 IVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFNDCGYPCPEHSNPFDFYMD 303
      :||:||||||| :| ||||: :|| ||::: |  :: :|  : || ||:: |  |: :|
Db      582 TLVLSIHQPRSNIFYLFDKLVLLSKGEMVYSGNAKKVSEFLRNEGYICPDNYNIADYLD 641
Qy      304 LT-SVDTQSKEREI----- 316
      :|      | | | |
Db      642 ITFEAGPQGKRRRIRNISDLEAGTDTNDIDNTIHQTTFTSSDGTQREWAHLAAHRDEIR 701
Qy      317 -----ETSKRVQMIESAYKKSACHKTLKNIERM----- 345
      |      ::  || |  :  : || :
Db      702 SLLRDEEDVEGTDGRRGATEIDLNTKLLHDKYKDSVYYAELSQEIEEVLSEGDEESNVLN 761
Qy      346 KHLKTLPMVPFKTKDSPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVL 405
      | |      : | | :| :| | :|: ||  ::  | ::| |||
Db      762 GDLPT-----GQQSAGFLQQLSILNSRSFKNMYRNPKLLLGNYLITLLSLFLGTLYY 814
Qy      406 RVRSNVLKGAIQDRVGLLY---QFVGATPYTGMLNAVNLFPPVLRVAVSDQESQDGLYQKWQ 462
      | || : |  |:|:| :  : |  :||:  : | : | : :| : |
Db      815 NV-SNDISG-FQNRMGLEFFILTIFYGFVTFGL----SSFALERIIFIKERSNNYYSP-- 866
Qy      463 MMLAYAL-----HVLPPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLT 517
      ||| :      |:| || :: | : |  ||: :  | :  :| :| |
Db      867 --LAYYISKIMSEVVPLRVVPPILLSLIVYPMTGLNMKDNF-FKCIGILILFNLGISLE 923
Qy      518 LVLLGIV---QNPNIVNSVVALLSIAGVLVSGFRLNIQEMP-IPFKIISYFTFQKYCSE 573
      :: :||:  | :|: ||: ||  | |: ||  | : :  : || : | : | |
Db      924 ILTIGIIFEDLNNSIILSVLVLL---GSLLFSGLFINTKNITNVAFKYLKNFSVFYYAYE 980
Qy      574 ILVVNEF-----YGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRETMNF 624
      |:|:|      ||||      |||| | |
Db      981 SLLINEVKTLMLKERKYGLNI-----EVPGATILSTFGF 1014
Qy      625 LILYSFIPALVILGI--VVFKIRDHL 648
      :: : :  : || :  ||| | :|
Db      1015 -VVQNLVFDIKILALFNVVFLIMGYL 1039

```

RESULT 5

T47648

ABC transporter-like protein - Arabidopsis thaliana

N;Alternate names: protein T15C9.80

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 19-May-2000

C;Accession: T47648

R;Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A;Reference number: Z24470

A;Accession: T47648

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-720 <MEW>

A;Cross-references: EMBL:AL132970

A;Experimental source: cultivar Columbia; BAC clone T15C9

C;Genetics:

A;Map position: 3  
A;Note: T15C9.80  
C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein  
F12L6.1; ATP-binding cassette homology

Query Match 18.5%; Score 614; DB 2; Length 720;  
Best Local Similarity 28.7%; Pred. No. 4.9e-38;  
Matches 182; Conservative 122; Mismatches 254; Indels 76; Gaps 17;

```

Qy      23 SSLEG--APATAPEPHSLGILHASYSVS-----HRVRPWWDITSCRQQWTRQILKDVS 73
      |||:| | | | : :|| | : || : : | : :| :|
Db      40 SSLDGDNDHLMRPVPFVLSFNNTYNVSVRRKLDHDLVPWRRTSFSK---TKTLLDNIS 96

Qy      74 LYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVL 133
      |:|: :||:|||||:|:|:|: | : : | | :|| ||: : :||:
Db      97 GETRDGEILAVLGASGSGKSTLIDALANRIAK-GSLKGTVTLNGEALQSRMLKVISAYVM 155

Qy     134 QSDTLLSSLTVRETLHYTALLAIRRGNPGSFQK-KVEAVMAELSLSHVADRIGNYSIGG 192
      | | | ||| ||| : | : | | | :|:|:| :| : | :||: |
Db     156 QDDLFPMLTVEETLMFAAEFRLPRSLPKSKKKLRVQALIDQLGIRNAAKTIIGDEGHRG 215

Qy     193 ISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQP 252
      || ||||| | : || | : |||:| | : | :| : | :|:| | |
Db     216 ISGGERRRVSIGIDIHDPVLFLEPTSGLDSTSAFMVVKVLKRIAESGSIIMSIIHQP 275

Qy     253 RSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTS----- 306
      : | |:: || | : | | : | | | | : | :|
Db     276 SHRVLSLDRLIFLSRGHTVFSGSPASLPSPFFAGFGNPIPENENQTEFALDLIRELEGS 335

Qy     307 -----VDTQSKEREIETSKRVQMIESAYKKSACHKTLK-----NIERMKHLK----- 349
      | : | :|:| | : : : | | : | | | : | | | :
Db     336 GGTRGLVEFNKKWQEMKKQSNPQTLTPASPSP--NLTLKEAISASISRGKLVSGGGGGS 393

Qy     350 -----TLPMPVPFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMGLFLL 401
      || : | : : : | | | | : | | : | | : | :| | :|
Db     394 SVINHGGGTIAVPAFANP----FWIEIKTLTRRSILNSRRQPELLGMRLATVIVTG-FIL 448

Qy     402 FFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQDGLYQKW 461
      | | | :|:|:| : | : | : :|:| : | : :| :|
Db     449 ATVFWRLDNSPKG-VQERLG-FFAFAMSTMFYTCADALPVFLQERYIFMRETAYNAYRRS 506

Qy     462 QMMLAYALHVLPPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFITLVLL 521
      :|:|: | : : : | : :| : | : :| | | |
Db     507 SYVLSHAIVTFPSLIFLSLAFVTTFWAVGLEGLMGFLFYCLIILASFWSGSSSFVTFLS 566

Qy     522 GIVQNPNIIV---NSVALLSIAGVLVGSGFLRNIQEMPIPFKIIISYFTFQKYCSEILVVN 578
      |:| |::| |||:| | | :| | | : | : | | | :|
Db     567 GVV--PHVMLGYTIVVAIL--AYFLLFSGFFINRDRI PQYWIWFHYLSLVKYPYEA VLQN 622

Qy     579 EFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKT 612
      || :| | :|:| : :
Db     623 EF-----SDPTECFVRGVQLFDNS 641

```

RESULT 6

C84423

probable ABC transporter [imported] - Arabidopsis thaliana



C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C;Accession: C84423  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;  
 Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell,  
 C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin,  
 L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams,  
 M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver,  
 G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser,  
 C.M.; Venter, J.C.  
 Nature 402, 761-768, 1999  
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis  
 thaliana.  
 A;Reference number: A84420; MUID:20083487; PMID:10617197  
 A;Accession: C84423  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-725 <STO>  
 A;Cross-references: GB:AE002093; NID:g4262239; PIDN:AAD14532.1; GSPDB:GN00139  
 C;Genetics:  
 A;Gene: At2g01320  
 A;Map position: 2

Query Match 18.4%; Score 610.5; DB 2; Length 725;  
 Best Local Similarity 29.7%; Pred. No. 9e-38;  
 Matches 166; Conservative 111; Mismatches 244; Indels 37; Gaps 11;

```

Qy      51 VRP----WWDITSC----RQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLDDAMSGR 102
      :||  | :||      : | :||:|  : |:: | | ||||| : :|:
Db      65 IRPVTIRWRNITCSLSDKSSKSVRFLKNVSGEAKPGRLLAIMGPSGSKTTLNVLAGQ 124

Qy     103 LGRAGT--FLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRR-G 159
      | :  | : ||| :  : ::  ::| | | | ||||| : | | :
Db     125 LSLSPRLHLSGLLEVNGKPSSSKAYK--LAFVRQEDLFFSQLTVRETLFAAELQLPEIS 182

Qy     160 NPGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDEP 219
      :  : | : : | |  || :| : || | : :| :| : | : ||
Db     183 SAEERDEYVNNLLKLGLVSCADSCVGDAKVRGISGGEKKRLSLACELIASPSVIFADEP 242

Qy     220 TTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPA- 278
      ||||| | : :: | :||:  | : ||||| : : || | :| : | :| : ||
Db     243 TTGLDAFQAEKVMETLQKLAQDGHTVICSIHQPRGSVYAKFDDIVLLTEGTLVYAG-PAG 301

Qy     279 -EMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAY--KKSAIC 335
      | | :| : | : |||| || :| || ||| | | : ||| : | : :| : :
Db     302 KEPLTYFGNFGFLCPEHVNPAEFLADLISVDYSSSETVYSSQKRVHALVDAFSQRSSSVL 361

Qy     336 HKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRLRRV-----TRNLVRNKLAVI 387
      : | : ::  |  : : | : : : ||:|  | || : :|
Db     362 YATPLSMKEETKNGMRPRRKAIVERTDGWWRQFFLLLKRAWMQASRDGPTNKVRARMSVA 421

Qy     388 TRLLQNLIMGLFLLFFVLVRVSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRA 447
      :  :  : :| : |  : :|||:| ||  :  | :|| ||
Db     422 SA-----VIFGSVFWRMGKSQTSIQDRMGLLQVAINTAMAALTKTGVGFPKERA 471

Qy     448 VSDQESQDGLYQKWQMMLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVARFGYFSAALL 507
      : | :|  | |  :| : :|  : :| :| |  | :| :||| |  :

```



```

Qy      308 DT-----QSKEREIETSKRVQMIE----SAYKKSACHKTLKNIERMKHLKTLP 352
      |          |:: | ||: | | : : | |:::
Db      338 GTRDLVEFNEKWQQNQATARATTQSRVSLKEAIAASVSRGKLVSGSSGANPISMETVSSYA 397

Qy      353 MVPFKTKDSPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVL 412
      |          :: :| :| :| :| : | : :| | | | : |
Db      398 NPP-----LAETFILAKRYIKNWIRTPELIGMRIGTVMVTGLLLATVYWRL-DNTP 447

Qy      413 KGAIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVL 472
      :|| |::| : | :| : : :| | : :| : | : :|| |
Db      448 RGA-QERMG-FFAFGMSTMFYCCADNIPVFIQERYIFLRETHNAYRTSSYVISHALVSL 505

Qy      473 PFSVVATMIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLTLVLLGIVQNPNI VNS 532
      | : :: |:: :||:| | : | | : : | : : |:: ||:: |
Db      506 PQLLALSIAFAATTFTVGLSGGLESFFYYCLIIYAAFWSGSSIVTFISGLI--PNVMS 563

Qy      533 -VALLSIAGVLVSGSGLRNIQEMPIPFKIISYFTFQKYCSEILVNEFYGLNFTCGSSN 591
      :| : :: | : || | :| : : | : || | : :|||
Db      564 YMVTIAYLSYCLLLGGFYINRDRIPLYWIWFHYISLLKYPYEAVLINEF----- 612

Qy      592 VSVTTNPMCAFTQGIQFIEKT 612
      :| | :|:| : |
Db      613 ----DDPSRCFVKGVQVFDGT 629

```

# RESULT 8

FYFFW

white protein - fruit fly (*Drosophila melanogaster*)

C;Species: *Drosophila melanogaster*

C;Date: 31-Dec-1990 #sequence\_revision 17-Feb-1995 #text\_change 19-Jan-2001

C;Accession: S08635; S07263; S10240

R;Pepling, M.; Mount, S.M.

Nucleic Acids Res. 18, 1633, 1990

A;Title: Sequence of a cDNA from the *Drosophila melanogaster* white gene.

A;Reference number: S08635; MUID:90221897; PMID:2109311

A;Accession: S08635

A;Molecule type: mRNA

A;Residues: 1-687 <PEP>

A;Cross-references: EMBL:X51749; NID:g8825; PIDN:CAA36038.1; PID:g8826

R;O'Hare, K.; Murphy, C.; Levis, R.; Rubin, G.M.

J. Mol. Biol. 180, 437-455, 1984

A;Title: DNA sequence of the white locus of *Drosophila melanogaster*.

A;Reference number: S07263; MUID:85134865; PMID:6084717

A;Accession: S07263

A;Molecule type: DNA

A;Residues: 1-24, 'LIFEIPYHCRVTAD', 30-

334, 'ITLHLNSYPWVPSVLPTTIRRTFTYRCWPLCPDGRSSPVIGSPRYG', 372-687 <OHA1>

A;Cross-references: EMBL:X02974

A;Experimental source: strain Canton S

R;O'Hare, K.

submitted to the EMBL Data Library, June 1985

A;Reference number: S10240

A;Accession: S10240

A;Molecule type: DNA

A;Residues: 1-24, 'LIFEIPYHCRVTAD', 30-687 <OHA2>

A;Cross-references: EMBL:X02974; NID:g10873; PIDN:CAA26716.1; PID:g10874

A;Experimental source: strain Canton S

C;Genetics:  
A;Gene: white; w  
A;Cross-references: FlyBase:FBgn0003996  
A;Introns: 24/3; 116/1; 334/2; 439/3; 483/3  
C;Superfamily: fruit fly white protein; ATP-binding cassette homology  
C;Keywords: ATP; glycoprotein; nucleotide binding; P-loop; transmembrane protein  
F;113-317/Domain: ATP-binding cassette homology <ABC>  
F;130-137/Region: nucleotide-binding motif A (P-loop)  
F;261-265/Region: nucleotide-binding motif B  
F;67,93,472,554,651/Binding site: carbohydrate (Asn) (covalent) #status  
predicted

Query Match 18.1%; Score 602.5; DB 1; Length 687;  
Best Local Similarity 28.8%; Pred. No. 3.3e-37;  
Matches 180; Conservative 131; Mismatches 220; Indels 95; Gaps 19;

```

Qy      66 RQILKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGR--LGRAGTFLGEVYVNGRALRRE 123
      : :||:|   |::: ::|||:|||||:|:: |   |   : ||: : :
Db      110 KHLKKNVCGVAYPGELLAVMGSSGAGKTTLNLAFRSPQGIQVSPSGMRLNLNGQPVDK 169

Qy      124 QFQDCFSYVLQSDTLSSLTVRETSLHYTALLAIRGNPGSFQK---KVEAVMAELSLSHV 180
      : |   :|| | | : ||| || | : | : |   ::::   :|: | : ||||
Db      170 EMQARCAIVQDDLFIGSLTAREHLIFQAM--VRMPRHLYRQVRARVDQVIQELSLSKC 227

Qy      181 ADRLIG-NYSLGGISTGERRRVSIAAQLLQDPKVMLEFDEPTTGLDCMTANQIVVLLVELA 239
      :||   : |:| |||:|:: |: | || :: | |||:| || | :| :| :|:
Db      228 QHTIIGVPGRVKGLSGGERKRLAFASEALTDPPLLICDEPTSGLDSTFAHSVVQVLKKLS 287

Qy      240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHSNPF 299
      :: : |:||||| |||:||||| :: : | : | |||:| :|||: | || : || |
Db      288 QKGKTVILTIHQPSSELFELFDKILLMAEGRVAF LGTPSEAVDFFSYVGAQCPTNYPAD 347

Qy      300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKS AICHKTLKNIERMKHLKTLPMVPFKTK 359
      ||: : :|   ||||: |: | : |   |   | :||:| | | : || |
Db      348 FYVQVLAV---VPGREIESRDRIAKICDNFAIS----KVAR DMEQLLATKNLE----KPL 396

Qy      360 DSP-----GVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFL-LFFVLRVRSN 410
      : |   | :   :|| | :::: | | ||:| : : : | |: : :
Db      397 EQPENGTYTKATWFMQFRAVLWRSWLSVLKEPLLVKVRLIQTTMVA ILIGLIFLGQQLTQ 456

Qy      411 VLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPPVLRVSDQESQDGLYQKWQMMLAYALH 470
      |   : : | :: |:   : :   :|:|   | :||: ||:   | :
Db      457 V---GVMNINGAIFLFLTNMTFQNVFATINVTSELPVFMREARSRLYRCDTYFLGKTIA 513

Qy      471 VLPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEF LTLVLLGIVQNP NIV 530
      || :   :||::: | :|| | |   | | | : :| | :
Db      514 ELPLFLT VPLVFTA IAYPMIGLRAGVLHF-----FNCLALVTLV--ANVS 556

Qy      531 NSVVALLSIAG-----VLVSGSGLRNIQEMPIPFKIISYFTFQKYCSEI 574
      |   |:| |   | : || | :|: ||:| :|: ||: :| :|
Db      557 TSFGYLISCASSSTSMALSVGPPVIIPFLFLGGFFLNSGSPVYLKWL SYLSWFRYANEG 616

Qy      575 LVVNEFYGL---NFTCGSSNVSVTTNPMCAFTQGIQFIEKTC P--GATS RFTMNFLLIYS 629
      |:|:|: :   :| |||   ||| |   |:| | :
Db      617 LLINQWADVEPGEISCTSSNT-----TCPSSGKVILETLNFSA--A 655

Qy      630 FIP----ALVILGIVVFKIRDHLISR 651

```

Db 656 DLPLDYVGLAIL-IVSFRVLAYLALR 680

RESULT 9

B96573

protein F12M16.17 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: B96573

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: B96573

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-590 <STO>

A;Cross-references: GB:AE005173; NID:g7769856; PIDN:AAF69534.1; GSPDB:GN00141

C;Genetics:

A;Gene: F12M16.17

A;Map position: 1

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.9%; Score 597; DB 2; Length 590;  
Best Local Similarity 29.6%; Pred. No. 7e-37;  
Matches 186; Conservative 113; Mismatches 270; Indels 60; Gaps 15;

Qy 29 PATAPEP-----HSLGILHASYSVSHRVRPWWDITS-CRQQWTRQILKDVS LYVESGQI 81  
| ||| : | : || : : : : ||||| | :|  
Db 4 PVKAPIPGGREISYRLETKNLSYRIGGNTPKFSNLCGLLSEKEEKVILKDVSCDARS AEI 63  
  
Qy 82 MCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSS 141  
| | ||:|||||: :::: | |:| ||| : : : :| | | |  
Db 64 TAIAGPSGAGKTTLLEILAGKVSH-GKVS GQVLVNGRPMDGPEYRRVSGFVPQEDALF PF 122  
  
Qy 142 LTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRV 201  
|||:||| |:||| : : | |: : : || | ||| || | ||| |||||  
Db 123 LTVQETLTYSALLRLKTKRKDA-AAKVRLIQELGLEHVADSRIGQGSRSISGGERRRV 181  
  
Qy 202 SIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELA-RRNRIVLTIHQPRSELFQLF 260  
|| :|: || |:| ||||:||| :| |:| || : : : :||| ||| : :

Db 182 SIGVELVHDPNVILIDEPTSGLDASALQVVTLLKDMTIKQGKTIVLTIHQPGFRILEQI 241

Qy 261 DKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMD----LTSVDTQSKEREI 316  
|:| :|| |:: |: : | | : : | | : ||| |||

Db 242 DRIVLLSNGMVVQNGSVYSLHQIKFSGHQIPRRVNVLEYAIDIAGSLEPIRTQSC-REI 300

Qy 317 ETSKRVMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKLGVLRRVT 376  
| | | | | : : : | | : : | |

Db 301 -----SCYGHS----KTKWSC----YISAGGELHQSDSHSNSVLEEVQILGQRSC 342

Qy 377 RNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGML 436  
:|: | | | | | | | | | | | | : | : :

Db 343 KNIFRTKQLFTTRALQASIAGLILGSIYLVN-GNQKKEAKVLRTG-FFAFILTFLLSSTT 400

Qy 437 NAVNLFPVLRAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEV 496  
: : | | : : | : : | | : | : : | : | : | :

Db 401 EGLPIFLQDRRILMRETSRRAYRVLSYVLADTLIFIPFLLIISMLFATPVYWLVLRLREL 460

Qy 497 ARFGYFSAALLAPHLIGEFLTLVLGLGIVQNPNIIVNSVVALLSIAGVLVSGFRLNIQEMP 556  
| ||| : | : : | | : ||: | : : ||: : |

Db 461 DGFLYFSLVIWIVLLMSNSFVACFSALVPNFIMGTSVISGL-MGSFFLFSGYFIAKDRI 519

Qy 557 IPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGA 616  
: : : | : || | | : | : : | : | : :

Db 520 VYWEFMHYLSLFKYPFECLMINEY-----RGDVFLKQQDLKE 556

Qy 617 TSRFTMNLILYSFIPALVILGIVVFKIR 645  
: : : | | : || : | : |

Db 557 SQKWS-NLGIMASFIVGYRVLGFFILWYR 584

# RESULT 10

T31958

hypothetical protein F02E11.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 31-Jan-2000

C;Accession: T31958

R;Favella, A.; Scheet, P.

submitted to the EMBL Data Library, July 1997

A;Description: The sequence of *C. elegans* cosmid F02E11.

A;Reference number: Z21104

A;Accession: T31958

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-658 <FAV>

A;Cross-references: EMBL:AF016661; PIDN:AAB66050.1; GSPDB:GN00020; CESP:F02E11.1

A;Experimental source: strain Bristol N2; clone F02E11

C;Genetics:

A;Gene: CESP:F02E11.1

A;Map position: 2

A;Introns: 115/3; 158/3; 214/3; 330/3; 368/2; 448/3; 525/1

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

Query Match 17.9%; Score 595.5; DB 2; Length 658;

Best Local Similarity 27.1%; Pred. No. 1.1e-36;

Matches 165; Conservative 121; Mismatches 255; Indels 67; Gaps 11;

Qy	72	VSLYVESGQIMCILGSSSGSKTTLDDAMSGRLGRAGTFGLGEVYVNGRALRREQFQDCFSY	131
Db	79	VSGVAEPGEVLALMGGSAGAKTTLMNILAHLDTNGVEYLGDVTVNGKKITKQKMRQMCA	138
Qy	132	VLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHVADRLIG-NYSL	190
Db	139	VQQVDLFCGTLTVREQLTYTAHMRMKNATVQQKMERNENVLDRMNLTDQCNTLIGIPNR	198
Qy	191	GGISTGERRRVSIQAQLLDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIH	250
Db	199	KGISIGEKKRLAFACEILTDPKILFCDEPTSGLDAFMASEVVRRALLDLANKGKTIIIVLH	258
Qy	251	QPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGC--YPCPEHSNPFDGYMDLTSVD	308
Db	259	QPSSTVFRMFHKVCFMATGKTVYHGAVDRLCPFFDKLGPDVRVPESYNPADFVMSEISI-	317
Qy	309	TQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKDSPGVFSKL	368
Db	318	--SPETEQedvtrieyliheyqnsdigTQMLK-----KTRTAVDEFEGGY	359
Qy	369	G-----VLLRRVTRNLVRNKLAVITRLLQNLI MGFLFFVLRVRS	409
Db	360	GDEDDGESRYNSTFTGTQFEILLKRSLRTTFRDPLLLRVRFQILATAILVGIVNWVRV-	418
Qy	410	NVLKG-AIQDRVGLLYQFVGATPYTGMLNAVNLFVPVLRASVDQESQDGLYQKWQMMLAYA	468
Db	419	--LKGPTIQNLEGVMYNCARDMTFLFYFPSNVNITSELPVFLREHKSNIYSVEAYFLAKS	476
Qy	469	LHVLPFSVVATMIFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEFRTLVLGIVQNPN	528
Db	477	LAELPQYTILPMIYGTHIYWMAGLVASVTSFLVFVFCITLTWVAVSIAYVGACIFGDEG	536
Qy	529	IVNSVVALLSIAGVLVSGSGLRNIEPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCG	588
Db	537	LVTTFMPMF-VLPMLVFGGFYVNANSIPVYYQYVSFVSWFKHGFEALEANQWKIDKISG	595
Qy	589	SSNVSVTTNPMCAFTQGFIQFIEKTCP-----GATSrftmnlilysfipalvil	637
Db	596	CDLI----NPLNATTGY-----CPASDGPGILTrrgidtPLYANVILlEMSffvyrii	645
Qy	638	GIVVFKIR	645
Db	646	GLVALKIR	653

A;Reference number: Z14679  
 A;Accession: T02567  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-740 <ROU>  
 A;Cross-references: EMBL:AC004697; NID:g3402671; PIDN:AAC28975.1; PID:g3402672  
 A;Experimental source: cultivar Columbia  
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.; Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.  
 submitted to the EMBL Data Library, July 1998  
 A;Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.  
 A;Reference number: Z14168  
 A;Accession: T00545  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-362 <ROW>  
 A;Cross-references: EMBL:AC004218; NID:g3355463; PIDN:AAC27826.1; PID:g3355464  
 A;Experimental source: cultivar Columbia  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.  
 Nature 402, 761-768, 1999  
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A;Reference number: A84420; MUID:20083487; PMID:10617197  
 A;Accession: C84816  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-740 <STO>  
 A;Cross-references: GB:AE002093; NID:g3402672; PIDN:AAC28975.1; GSPDB:GN00139  
 C;Genetics:  
 A;Gene: At2g39350; T16B24.1; F12L6.1  
 A;Map position: 2  
 C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-binding cassette homology  
 C;Keywords: ATP  
 F;110-310/Domain: ATP-binding cassette homology <ABC>

Query Match 17.8%; Score 591.5; DB 1; Length 740;  
 Best Local Similarity 27.5%; Pred. No. 2.5e-36;  
 Matches 191; Conservative 123; Mismatches 267; Indels 113; Gaps 19;

Qy	35	PHSLGILHASYSVSHRVRPWWD-----ITSCRQQWTRQILKDVSPLYVESGQ	80
		: : :            :   : :  :    :	
Db	64	PFVLSFDNLTYNVS--VRPKLDFRNLFPRRRTEDPEIAQTARPKTKTLLNNISGETRDGE	121
Qy	81	IMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLS	140
		: :        : :  :  :   :     :    : : :	
Db	122	IMAVLGASGSGKSTLIDALANRIAK-GSLKGTVKLNGETLQSRMLKVISAYVMQDDLFP	180
Qy	141	SLTVRETLHYTALLAIRRGNPGSFQK-KVEAVMAELSLSHVADRLIGNYSLGGISTGERR	199
		:   :     :  : :  :  : :   : :	



```

Db      181 MLTVEETLMFAAEFRLPRSLPKSKKKLRVQALIDQLGIRNAAKTIIGDEGHRGISGGERR 240
Qy      200 RVSIAAQLLQDPKVMFLFDEPTTGLDCMTANQIVLLVELARRNRIVVLTIHQPRSELFQL 259
      |||| :: || :: ||||:|||| :| :| :| :| :| :| :| :| :|
Db      241 RVSIGIDIIHDPILLFLDEPTSGLDSTSAFMVVKVLKRIAQSGSIVIMSIHQPSHRVLGL 300
Qy      260 FDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPFDFYMDLTSVDTQSKERE---- 315
      |:: || | :: |::| : || : | | ||: | :| :|| :| |
Db      301 LDRLIFLSRGHTVYSGSPASLPRFFTEFGSPIPENENRTEFALDLI-----RELEGSAG 354
Qy      316 -----IETSKRVQMIE-----SAYKKSACHKTLKNIERMK----- 346
      || :|: | :: | | : :| | |
Db      355 GTRGLIEFNKKWQEMKKQSNRQPPLTPSSPYPNLTLKEAIAASISRGKLVSGGESVAHG 414
Qy      347 ----HLKTLPMVPFKTKDSPGVFSKLGVLLRRVTRNLVRNKLAVITRLLQNLMGLFLLF 402
      : || : | :: :: | :| | | | : :| | | :|
Db      415 GATTNTTTLAVPAFANP---MWIEIKTLSKRSMNSRRQPELFGIRIASVVITG-FILA 469
Qy      403 FVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQ 462
      | | | | :|:| :| :| : :| :| :| :| :| :| :|
Db      470 TVFWRLDNSPKG-VQERLG-FFAFAMSTMFYTCADALPVFLQERYIFMRETAYNAYRRSS 527
Qy      463 MMLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLTLVLLG 522
      :|::|: | : :: |:: || :|| : :: :|| | | |
Db      528 YVLSHAIVSFPSLIFLSVAFAATTYWAVGLDGGTLGLLFYCLIIILASFWSGSSEFVTFLSG 587
Qy      523 IVQNPNIIV---NSVVALLSIAGVLVSGSGLRNIQEMPIPKIISYFTFQKYCSEILVVNE 579
      :| |::: |||:| | :| || | :| : | : || | :| ||
Db      588 VV--PSVMLGYTIVVAIL--AYFLLFSGFFINRNRIPDYWIWFHYMSLVKYPYEAVLQNE 643
Qy      580 FYGLN--FTCG-----SSNVSVTTNPMCAFTQG----- 605
      | | | | | :| :| : | |
Db      644 FSDATKCFVRGVQIFDNTPLGELPEVMKLLKLTGTVSKSLGVTISSTTCLTTGSDILRQQG 703
Qy      606 -IQFIEKTCPGATSRFTMNFLILYSFIPALVILG 638
      :| : | | | | ||: | :||
Db      704 VVQLSKWNCLFITVAFGFFFRILFYF---TLLLG 734

```

RESULT 12

G02068

white homolog - human

C;Species: Homo sapiens (man)

C;Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 02-Feb-2001

C;Accession: G02068

R;Croop, J.M.; Tiller, G.; Fletcher, J.A.; Lux, M.; Raab, E.; Goldenson, D.; Arciniegas, S.; Son, D.; Wu, R.

submitted to the EMBL Data Library, August 1995

A;Reference number: H00769

A;Accession: G02068

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-638 <CRO>

A;Cross-references: EMBL:U34919; NID:g1314276; PIDN:AAC51098.1; PID:g1314277

C;Genetics:

A;Gene: white

C;Superfamily: fruit fly white protein; ATP-binding cassette homology

C;Keywords: ATP; nucleotide binding; P-loop  
F;61-253/Domain: ATP-binding cassette homology <ABC>  
F;78-85/Region: nucleotide-binding motif A (P-loop)

```

Query Match      17.8%; Score 590.5; DB 2; Length 638;
Best Local Similarity 26.3%; Pred. No. 2.4e-36;
Matches 164; Conservative 142; Mismatches 266; Indels 51; Gaps 14;

Qy      44 SYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRL 103
      |||  |||  |:: : :|| :|  ||::: |:| ||:|:|:|: :|
Db      43 SYSVPE--GPWW-----RKKGYKTLLKGISGKFNSGELVAIMGPSGAGKSTLMNIIAGY- 94

Qy     104 GRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGS 163
      |  |  | :||  |:  |::| ||  ||:| : :| | : : |
Db      95 -RETGMKGAVLINGLPRDLRCFRKVCYIMQDDMLLPHLTVQEAAMVSAHLKLQEKDEGR 153

Qy     164 FQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGL 223
      :: |: :: | |  |  |  | :| |:|:|:|:| |: :| || |||:|:|
Db     154 -REMVKEILTALGLLSCA-----NTRTGSLSGGQRKRLAIALELVNPPVMFFDEPTSGL 207

Qy     224 DCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDF 283
      | :  |:| |:  ||: | : : |||| :|:|:|:|: :| | : : |  : :
Db     208 DSASCFQVVSIMKGLAQGGRSIICTIHQPSAKLFELFDQLYVLSQGQCVRGKVCNLVPY 267

Qy     284 FNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS-----I 334
      | |  || : || || |:: | :  :  : : | | :| :| :
Db     268 LRDLGLNCPTYHNPADVFMEVASGEYGDQNSRLVRVAVREGMCDSDHKRDLGGDAEVNPF 327

Qy     335 CHKTLKNIERMKHLKTLPMVPFKTKDSPGV-----FSKLGVLLRRVTRNLVRNKL 384
      |:  : :: : | || |  ||| :  : : | :| :| : :| : :
Db     328 WHRPSEEVKQTKRLKGL-----RKDSSSMEGCHSFSASCLTQFCILFKRTFLSIMRDSV 381

Qy     385 AVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLF 444
      |:  : :|:| :  | : : | :  | |: :  : : : | ||:
Db     382 LTHLRITSHIGIGLLIGLLYLIGIGNEAKK--VLSNSGFLFFSMLFLMFAALMPTVLT 439

Qy     445 LRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVAREGYFSA 504
      | :| :  |  || : :|| : : : |:| |  :  | |:|
Db     440 EMGVFLREHLNYWYLSLKAYYLAKTMADVPFQIMFPVAYCSIVYWMTSQPSDAVAVLFAA 499

Qy     505 ALLAPHLIGEFLLTLVLLGIVQNPNIIVNSVVALLSIAGVLVGSGLFRNIQEMPIPFKIISY 564
      |: : | | |:|  | : | : : ||: ||  : :| : :|
Db     500 LGTMTSLVAQSLGL-LIGAASLSLQVATFVGPVTAIPVLLFSGFFVSFDTIPTYLQWMSY 558

Qy     565 FTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF 624
      : : :| | : : : |||:  : :  : | | | : | :  : : :|
Db     559 ISYVRYGFEVILS-IYGLD---REDLHCDIDETCHF-QKSEAILRELDVENAKLYLDF 612

Qy     625 LILYSFIPALVILGIVV--FKIR 645
      : :| | :| : :  | :||
Db     613 IVLGIFFISLRLLIAYFVLRYSKIR 635

```

RESULT 13  
C86441  
probable ABC transporter [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: C86441

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: C86441

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-646 <STO>

A;Cross-references: GB:AE005172; NID:g11136734; PIDN:AAG31315.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-binding cassette homology

Query Match 17.7%; Score 589.5; DB 2; Length 646;

Best Local Similarity 29.9%; Pred. No. 2.9e-36;

Matches 183; Conservative 112; Mismatches 246; Indels 71; Gaps 22;

```
Qy      10 GGSM--GLQVNRGSQS-SLEGAPATAPEPHSLGILHASYSVS-----HRVRPWWDITSCR 61
      || | ||      :|| | : |      :| | : : | ::      ::| :      : |
Db      14 GGVMVQGLPDMSDTQSKSVLAFPTITSQP---GLQMSYPITLKEVVYKVK-IEQTSQCM 69

Qy      62 QQW---TRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGR 118
      |      : ||      :: | | : : : || ||||| ||| | : || | : ||
Db      70 GSWKSKEKTILNGITGMVCPGEFLAMLGPGSGSGKTTLLSALGGRLSK--TFSGKVMYNG- 126

Qy      119 ALRREQFQDCF---SYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQK-----KVE 169
      : | |      : | | | | ||| ||| : ||| :      | | :      | :
Db      127 ----QPFSGCIKRRTGFVAQDDVLYPHLTWVETLFFTALLRL----PSSLTRDEKAEHVD 178

Qy      170 AVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGLDCMTAN 229
      | : ||| | :      : : ||      ||| ||| : ||| : : | : ||| : ||| || :
Db      179 RVIAELGLNRCTNSMIGGPLFRGISGGEKRVSIGQEMLINPSLLLLDEPTSGLDSTTAH 238

Qy      230 QIVVLLVELARRNRIVVLTIIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND CGY 289
      : || : ||      | || |||| | : : : ||| : : || | | : | : : : | :
Db      239 RIVTTIKRLASGGRTVVTTIIHQPSRIYHMFDKVLLSEGPSPIYYGAASSAVEYFSSLGF 298

Qy      290 PCPEHSNPFDFYMDLTS---VDTQSKEREIETSKRVQMIESAYK-----KSAICHKTL 339
      || | : || :      ||| : | |      : : ||| :      | : : | :
```

Db 299 STSLTVNPADLLLDLANGIPDQTQKETSEQEQKTVKETLVSAYEKNISTKLKAELCNAES 358  
 Qy 340 KNIERMK-HLKTLPMPVPFKTKDSPGVFSKLGVLRRVTRNL---VRNKLAVITRLLQNLI 395  
 : | | | : | : : |||: | | : : :  
 Db 359 HSYEYTKAAAKNLKSEQWCTT---WWYQFTVLLQRGVRERRFESFNKLRIQVISVAFL 414  
 Qy 396 MGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVAVSDQESQD 455  
 | ||: : | ||| ||: | : : || | : : :  
 Db 415 GG--LLWW-----HTPKSHIQDRTALLFFFSVFWGFYPLYNVAVFTFPQEKRLIKERS 466  
 Qy 456 GLYQKWQMMLAYALHVLPPFSVVATMIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEF 515  
 |:|: :| : || : | : || || |: | :| |: :  
 Db 467 GMYRLSSYFMARNVGDLPLELALPTAFVFIIYWMGGLKPDPTTFILSLVLVLYSVLVAQG 526  
 Qy 516 LTLVLLGIVQNPNIIVNSVVALLSIAGVLVGSGFLRNIQEMPIPFKI-ISYFTFQKYCSEI 574  
 | | : : | : : : : : | : : :| | | : : | : : || : :  
 Db 527 LGLAFGALLMNIKQATTLASVTTLVFLIAGGYY--VQQIP-PFIVWLKYLSSYSYCYKL 582  
 Qy 575 LVVNEFYGLNFT 586  
 |: |: :|  
 Db 583 LL-----GIQYT 589

# RESULT 14

G84791

probable ABC transporter [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001

C;Accession: G84791

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;  
 Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell,  
 C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin,  
 L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams,  
 M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver,  
 G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser,  
 C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: G84791

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-755 <STO>

A;Cross-references: GB:AE002093; NID:g4056489; PIDN:AAC98055.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g37360

A;Map position: 2

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein  
 F12L6.1; ATP-binding cassette homology

Query Match 17.6%; Score 584; DB 2; Length 755;

Best Local Similarity 27.2%; Pred. No. 9.4e-36;

Matches 173; Conservative 128; Mismatches 253; Indels 82; Gaps 19;

Qy 21 SQSSLEGAPAT--APEPHSLGILHASYSVSHRVRPWWDITSCRQQW-----TRQILKDV 72  
 | :| |||: : | | :||| : : : : ||: | :| :

Db 79 SFNSWASAPASSISSSPFVLSFTDLTYSVKIQ-KKFNPLACCRRSGNDSSVNTKILLNGI 137

Qy 73 SLYVESGQIMCILGSSSGSKTTLTDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYV 132  
| :| :||:||||:|:|:| : : : | : :|| | : :||

Db 138 SGEAREGEMMAVLGASGSGKSTLIDALANRIAK-DSLGRSITLNGEVLESSMQKVISAYV 196

Qy 133 LQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKK-----VEAVMAELSLSHVADRLIGN 187  
:| | | ||| ||| :| : | | || | :| :| | :||:

Db 197 MQDDLLFPMLTVEETLMFSAEFRL----PRSLSKKKKKARVQALIDQLGLRSAAKTVIGD 252

Qy 188 YSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGLDCMTANQIVVLLVELARRNRIVVL 247  
|:| ||||| ||| :| || :| ||||:| | :| :| :| :| :

Db 253 EHRGVSGGERRRVSIGNDIIHDPILFLDEPTSGLDSTSAYMVIKVLQRIAQSGSIVIM 312

Qy 248 TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTSV 307  
:|||| : | |:| ||| :| |:| : |:| :| |:| :| :| :

Db 313 SIHQPSYRIMGLLDQLIFLSKGN TVYSGSPTHLPQFFSEFKHIPENENKTEFALDLI-- 370

Qy 308 DTQSKEREIETSKRVQMIE-----SAYKKSACHKTLKNIERMKHLKTLTP----- 352  
:| | | :| :| :| :| :| :| :| :|

Db 371 ----RELEYSTEGTKPLVEFHQWRAKQAPSNNN---NKRNTNVSSLKEAITASISRGK 423

Qy 353 -----MVP-FKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLF 399  
: | |:| :| : :| :| | | : | | :| :

Db 424 LVSGATNNNSSNLTPSFQTFANP-FWIEMIVIGKRAILNSRRQPELLGMRLGAVMVTGII 482

Qy 400 LLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQDGLYQ 459  
| : | ||| |:|:| :| :| :| :| :| :| :| :

Db 483 LATMFTNL-DNSPKGA-QERLG-FFAFAMSTTFYTCAEAI PVFLQERYIFMRETAYNAYR 539

Qy 460 KWQMMLAYALHVLFPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLTV 519  
: :| :| :| :| :| :| :| :| :| :| :| :

Db 540 RSSYVLSQSIISIPALIVLSASFAATTFWAVGLDGGANGFFFFFYFTILASFWAGSSFVTF 599

Qy 520 LLGIVQNPNIIV---NSVVALLSIAGVLVGSGFLRNIQEMPPIPKIISYFTFQKYCSEILV 576  
| |:| :| :| :| :| :| :| :| :| :| :| :| :

Db 600 LSGVI--PNVMLGFTVVVAIL--AYFLLFSGFFISRDRIPVYWLWFHYISLVKYPYEGVL 655

Qy 577 VNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKT 612  
||| || | :|:| : :

Db 656 QNEF-----QNPTRCFARGVQLFDNS 676

# RESULT 15

D96553

hypothetical protein F5D21.6 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C;Accession: D96553

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.; Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.; Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.; Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda, M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken, S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: D96553

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-687 <STO>

A;Cross-references: GB:AE005173; NID:g10092349; PIDN:AAG12758.1; GSPDB:GN00141

C;Genetics:

A;Gene: F5D21.6

A;Map position: 1

C;Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-binding cassette homology

Query Match 17.5%; Score 583.5; DB 2; Length 687;  
Best Local Similarity 28.0%; Pred. No. 9e-36;  
Matches 167; Conservative 117; Mismatches 239; Indels 73; Gaps 16;

Qy 33 PEPHSLGILHASYSVSHRVRPWWITSCRQW----TRQILKDVSLYVESGQIMCILGSS 88  
| | :| :| | | :| :| :| :| :| :| :| :|

Db 13 PPPAEIG--RGAYLA-----WEDLTVVIPNFSGGPTRRLDGLNGHAEPGRIMAIMGPS 64

Qy 89 GSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETL 148  
| | | :| | | :| | | | :| :| :| :| :| :| :| :|

Db 65 GSGKSTLLDSLGRRLARNVIMTGNLLLNKKARLD--YGLVAVVTQEDILMGTLTVRETI 122

Qy 149 HYTALLAIRRGNPGSFQKK-----VEAVMAELSLSHVADRLIGNYSLGGISTGERRRVS 203  
| :| | : | :| | | :| | | | | :| | | :| | :|

Db 123 TYSAHLRL---SSDLTKKEEVNDIVEGTIIELGLQDCADRVIGNWHSRGVSGGERKRVSV 178

Qy 204 AAQLLQDPKVMLEFDEPTTGLDCMTANQIVVLLVELAR-RNRIVVLTIHQPRSELFQLFDK 262  
| :| | :| :| | | :| :| | :| | | :| | | :| |

Db 179 ALEILTRPQILFLDEPTSGLDASAFFVIQALRNIARDGGRTVVSIIHQPSSEVFALFDD 238

Qy 263 IAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRV 322  
: :| | :| :| :| :| :| :| :| :| :| :| :| :|

Db 239 LFLSSGETVYFGESKFAVEFFAEAGFPCPKRNPSDHFLRCINSDFDTVTATLKGSQRI 298

Qy 323 -----QMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKDSP 362  
: :| :| :| :| :| :| :| :| :| :| :| :| :|

Db 299 RETPATSDPIMNLATSEIKARLVEN-YRRSVYAKSAKSRIELASIEGHHGMEVRKGSEA 357

Qy 363 GVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRVGL 422  
| :| | :| | :| :| :| :| :| :| :| :| :|

Db 358 TWFKQLRTLTKRSFVNMCRD-----IGYYWSRIVYIVVSFCVGTIFYDVGH 404

Qy 423 LYQFVGATP-----YTGMLNAVNL--FPVL---RAVSDQESQDGLYQKWQMLLAYALHVL 472  
| :| | :| :| :| :| :| :| :| :| :| :|

Db	405	SYTSILARVSCGGFITGFMTFMSIGGFPSFIEEMKVIFYKERLSGYYGVSVYIISNYVSSF	464
Qy	473	PFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIENS	532
		:   :   :    : : :  : :    :  : :     :	
Db	465	PFLVAIALITGSITYNMVKFRPGVSHWAFFCLNIFFSVSVIESLMMVVASLV--PNFLMG	522
Qy	533	VVALLSIAG-VLVGSGFLRNIQEMPIPF--KIISYFTFQKYCSEILVVNEFYGLNF	585
		::     ::       : ::      : :: : :   :	
Db	523	LITGAGIIGIIMMTSGFFRLLPDLPKVFWRYPISEMSYGSWAIQGAYKNDFLGLF	578

Search completed: February 27, 2004, 07:18:55  
Job time : 16.5049 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 07:17:39 ; Search time 29.2557 Seconds  
(without alignments)  
4698.604 Million cell updates/sec

Title: US-09-989-981A-6  
Perfect score: 3326  
Sequence: 1 MGDLSILTPGSGMGLQVNRG.....PALVILGIVVFKIRDHLISR 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result	Query	
No.	Score	Match Length DB ID Description



1	3326	100.0	651	9	US-09-837-992-3	Sequence 3, Appli
2	3326	100.0	651	10	US-09-989-981A-6	Sequence 6, Appli
3	3326	100.0	651	14	US-10-090-455-6	Sequence 6, Appli
4	2744.5	82.5	652	9	US-09-837-992-1	Sequence 1, Appli
5	2744.5	82.5	652	10	US-09-989-981A-2	Sequence 2, Appli
6	1308	39.3	256	15	US-10-104-047-2795	Sequence 2795, Ap
7	697	21.0	672	10	US-09-989-981A-4	Sequence 4, Appli
8	697	21.0	673	10	US-09-989-981A-8	Sequence 8, Appli
9	697	21.0	673	14	US-10-090-455-7	Sequence 7, Appli
10	682.5	20.5	655	10	US-09-961-086-1	Sequence 1, Appli
11	682.5	20.5	655	15	US-10-405-806-13	Sequence 13, Appl
12	680.5	20.5	655	9	US-09-981-353-35	Sequence 35, Appl
13	680.5	20.5	655	14	US-10-120-687-61	Sequence 61, Appl
14	680.5	20.5	655	15	US-10-405-806-2	Sequence 2, Appli
15	674.5	20.3	655	9	US-09-866-866A-10	Sequence 10, Appl
16	674.5	20.3	655	14	US-10-090-455-5	Sequence 5, Appli
17	672.5	20.2	655	9	US-09-866-866A-27	Sequence 27, Appl
18	660	19.8	657	9	US-09-866-866A-14	Sequence 14, Appl
19	627	18.9	1095	15	US-10-369-493-2025	Sequence 2025, Ap
20	621	18.7	1049	15	US-10-369-493-1520	Sequence 1520, Ap
21	602.5	18.1	663	13	US-10-108-605-245	Sequence 245, App
22	598.5	18.0	674	14	US-10-090-455-4	Sequence 4, Appli
23	598.5	18.0	674	16	US-10-429-160-10	Sequence 10, Appl
24	595.5	17.9	658	15	US-10-369-493-5347	Sequence 5347, Ap
25	590.5	17.8	638	13	US-10-072-621-10	Sequence 10, Appl
26	585.5	17.6	646	13	US-10-072-621-9	Sequence 9, Appli
27	585.5	17.6	646	14	US-10-090-455-2	Sequence 2, Appli
28	578.5	17.4	627	14	US-10-090-455-8	Sequence 8, Appli
29	578	17.4	604	9	US-09-745-763-197	Sequence 197, App
30	571.5	17.2	646	14	US-10-079-087-2	Sequence 2, Appli
31	570.5	17.2	646	13	US-10-154-452-4	Sequence 4, Appli
32	567.5	17.1	646	14	US-10-090-455-13	Sequence 13, Appl
33	565.5	17.0	599	15	US-10-210-130-14	Sequence 14, Appl
34	562.5	16.9	646	13	US-10-154-452-8	Sequence 8, Appli
35	554	16.7	559	15	US-10-369-493-5740	Sequence 5740, Ap
36	545.5	16.4	608	15	US-10-369-493-5748	Sequence 5748, Ap
37	540.5	16.3	676	15	US-10-369-493-3799	Sequence 3799, Ap
38	517.5	15.6	610	15	US-10-369-493-5687	Sequence 5687, Ap
39	517.5	15.6	639	15	US-10-369-493-6184	Sequence 6184, Ap
40	504	15.2	695	15	US-10-369-493-6199	Sequence 6199, Ap
41	496.5	14.9	560	15	US-10-369-493-12899	Sequence 12899, A
42	496	14.9	551	15	US-10-369-493-3562	Sequence 3562, Ap
43	485	14.6	545	14	US-10-083-357-1335	Sequence 1335, Ap
44	430.5	12.9	1395	15	US-10-369-493-4054	Sequence 4054, Ap
45	425.5	12.8	615	10	US-09-949-029-24	Sequence 24, Appl

#### ALIGNMENTS

RESULT 1  
 US-09-837-992-3  
 ; Sequence 3, Application US/09837992  
 ; Patent No. US20020081687A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tian, Hui

```
; APPLICANT:  Schultz, Joshua
; APPLICANT:  Shan, Bei
; APPLICANT:  Tularik Inc.
; TITLE OF INVENTION:  Sitosterolemia Susceptibility Gene (SSG):  Compositions
; TITLE OF INVENTION:  and Methods of Use
; FILE REFERENCE:  018781-006020US
; CURRENT APPLICATION NUMBER:  US/09/837,992
; CURRENT FILING DATE:  2001-04-18
; PRIOR APPLICATION NUMBER:  US 60/198,465
; PRIOR FILING DATE:  2000-04-18
; PRIOR APPLICATION NUMBER:  US 60/204,234
; PRIOR FILING DATE:  2000-05-15
; NUMBER OF SEQ ID NOS:  45
; SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH:  651
; TYPE:  PRT
; ORGANISM:  Homo sapiens
; FEATURE:
; OTHER INFORMATION:  human sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION:  amino acid sequence
US-09-837-992-3
```

```
Query Match          100.0%;  Score 3326;  DB 9;  Length 651;
Best Local Similarity 100.0%;  Pred. No. 1.1e-309;
Matches 651;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

```
Qy      1 MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC 60
|
Db      1 MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC 60

Qy     61 RQQWTRQILKDVSPLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL 120
|
Db     61 RQQWTRQILKDVSPLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL 120

Qy    121 RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV 180
|
Db    121 RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV 180

Qy    181 ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR 240
|
Db    181 ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR 240

Qy    241 RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGPCPEHSNPFDF 300
|
Db    241 RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGPCPEHSNPFDF 300

Qy    301 YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD 360
|
Db    301 YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD 360

Qy    361 SPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV 420
|
Db    361 SPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV 420

Qy    421 GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMMLAYALHVLFPFSVVATM 480
|
```

Db 421 GLLYQFVGATPYTGMLNAVNLFVFLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM 480

Qy 481 IFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLTLVLLGIVQNPNIIVNSVVALLSIA 540  
 |||

Db 481 IFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLTLVLLGIVQNPNIIVNSVVALLSIA 540

Qy 541 GVLVSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC 600  
 |||

Db 541 GVLVSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC 600

Qy 601 AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR 651  
 |||

Db 601 AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR 651

RESULT 2

US-09-989-981A-6

; Sequence 6, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

; FILE REFERENCE: 018781-007320US

; CURRENT APPLICATION NUMBER: US/09/989,981A

; CURRENT FILING DATE: 2002-07-23

; PRIOR APPLICATION NUMBER: US 60/252,235

; PRIOR FILING DATE: 2000-11-20

; PRIOR APPLICATION NUMBER: US 60/253,645

; PRIOR FILING DATE: 2000-11-28

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 651

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: human ABCG5 (hABCG5)

US-09-989-981A-6

Query Match 100.0%; Score 3326; DB 10; Length 651;

Best Local Similarity 100.0%; Pred. No. 1.1e-309;

Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC 60  
 |||

Db 1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC 60

Qy 61 RQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRAL 120  
 |||

Db 61 RQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRAL 120

Qy 121 RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV 180

Db	121	RREQFQDCFSYVLQSDTLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFD	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFD	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATM	480
Qy	481	IFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINIVNSVVALLSIA	540
Db	481	IFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPINIVNSVVALLSIA	540
Qy	541	GVLVSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMNLILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMNLILYSFIPALVILGIVVFKIRDHLISR	651

RESULT 3

US-10-090-455-6

; Sequence 6, Application US/10090455

; Publication No. US20030027259A1

; GENERAL INFORMATION:

; APPLICANT: Chen, Hongyun

; APPLICANT: Le Bihan, Stephane

; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF

; FILE REFERENCE: 100103.406

; CURRENT APPLICATION NUMBER: US/10/090,455

; CURRENT FILING DATE: 2002-03-01

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 651

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-090-455-6

Query Match 100.0%; Score 3326; DB 14; Length 651;  
Best Local Similarity 100.0%; Pred. No. 1.1e-309;

Matches 651; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Db	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSC	60
Qy	61	RQQWTRQILKDVSPLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Db	61	RQQWTRQILKDVSPLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Qy	481	IFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Db	481	IFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMFLILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMFLILYSFIPALVILGIVVFKIRDHLISR	651

#### RESULT 4

US-09-837-992-1

; Sequence 1, Application US/09837992

; Patent No. US20020081687A1

; GENERAL INFORMATION:

; APPLICANT: Tian, Hui

; APPLICANT: Schultz, Joshua

; APPLICANT: Shan, Bei

; APPLICANT: Tularik Inc.

; TITLE OF INVENTION: Sitosterolemia Susceptibility Gene (SSG): Compositions

```
; TITLE OF INVENTION:  and Methods of Use
; FILE REFERENCE: 018781-006020US
; CURRENT APPLICATION NUMBER: US/09/837,992
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: US 60/198,465
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 60/204,234
; PRIOR FILING DATE: 2000-05-15
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 652
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse sitosterolemia susceptibility gene (SSG)
; OTHER INFORMATION: amino acid sequence
US-09-837-992-1
```

```
Query Match          82.5%; Score 2744.5; DB 9; Length 652;
Best Local Similarity 80.2%; Pred. No. 6.6e-254;
Matches 523; Conservative 64; Mismatches 64; Indels 1; Gaps 1;
```

```
Qy      1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDDITS 59
      ||:| |:| |:| :||| ||| | | |||:| | |||:| | |||:| |
Db      1 MGELPFLSPEGARGPHINRGSLSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60

Qy      60 CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119
      |:|:| |||||:|||||:|||||:|||||:||||| | || |||:| |
Db      61 CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVFVNGCE 120

Qy      120 LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH 179
      |||:||||| ||||| |||:|:| : : ||||| |||||
Db      121 LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSH 180

Qy      180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLDEPTTGLDCMTANQIVVLLVELA 239
      |||:|:|:| |||:|||||:|||||: |||||: |||||: || |||
Db      181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240

Qy      240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPFD 299
      ||:|||:|:||||| |||||:|:|:| || | |||:|||||
Db      241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCGYPCEHSNPFD 300

Qy      300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK 359
      |||||:||||| |||||:| |:|:| | || |:||| :|:|||||
Db      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360

Qy      360 DSPGVFVSKLGVLLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419
      | ||:| |||||:||||| || |||:|||||:|:|:|:| |||:| |
Db      361 DPPGMFGKLGVLRLRRVTRNLVRNKLAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR 420

Qy      420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLVLPFSVAT 479
      ||||| |||||:||||| |||||:| | |||||:| |
Db      421 VGLLYQLVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYHKWQMLLAYVLHVLVLPFSVIAT 480

Qy      480 MIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPINIVNSVALLSI 539
      :|||||:|||||:|||||:|||||:|||||:|||||
```



```

      |||::||:|: ||||:|||||:|||||:|||||:|||||:|||||:
Db      181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240
Qy      240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPF 299
      ||:||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      241 RRDRIVIVTIHQPRSELFQHFDKIAILTYGELVFCGTPEEMLGFFNNGYPCPEHSNPF 300
Qy      300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK 359
      |||||:|||||:|||||:| |:|:| | | |:||| ::|||||
Db      301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPMPFKTK 360
Qy      360 DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419
      | ||:| |||||:|||| ||| ||:|||||:|::||:| ||||:|
Db      361 DPPGMFGKLGVLRRVTRNLMRNKQAVIMRLVQNLIMGLFLIFYLLRVQNNTLKGAVQDR 420
Qy      420 VGLLYQFVGATPYTGMLNAVNLFPPVLRVSDQESQDGLYQKWQMLLAYALHVLPPFSVAT 479
      ||||| |||||:|||||:|||||:|||||:|||||:|
Db      421 VGLLYQLVGATPYTGMLNAVNLFPMRLAVSDQESQDGLYHKWQMLLAYVLHVLPPFSVIAT 480
Qy      480 MIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPINVSIVALLSI 539
      :|||||:|||||:|||||:|||||:|||||:|||||:
Db      481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPINVSIVALLSI 540
Qy      540 AGVLVGSGLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599
      :|:|:||||:||||| ||: ||||| ||||| ||||| || | |: :|
Db      541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSMLNHPM 600
Qy      600 CAFTQGIQFIEKTCPGATSRTMNFILILYSFIPALVILGIVVFKIRDHLISR 651
      || |||:|||||:||||| ||||| |||||:|::||:|
Db      601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYLISR 652

```

RESULT 6

US-10-104-047-2795

; Sequence 2795, Application US/10104047

; Publication No. US20030236392A1

; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA

; FILE REFERENCE: H1-A0105

; CURRENT APPLICATION NUMBER: US/10/104,047

; CURRENT FILING DATE: 2002-03-25

; PRIOR APPLICATION NUMBER:

; PRIOR FILING DATE:

; NUMBER OF SEQ ID NOS: 4096

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2795

; LENGTH: 256

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-104-047-2795

Query Match 39.3%; Score 1308; DB 15; Length 256;

Best Local Similarity 100.0%; Pred. No. 9.2e-117;

Matches 256; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 396 MGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPPVLRVSDQESQD 455



```

Db          1 MGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQD 60
Qy          456 GLYQKWQMMLAYALHVLPPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEF 515
Db          61 GLYQKWQMMLAYALHVLPPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEF 120
Qy          516 LTLVLLGIVQNPNIIVNSVVALLSIAGVLVGSGLRNIQEMPIPFKIISYFTFQKYCSEIL 575
Db          121 LTLVLLGIVQNPNIIVNSVVALLSIAGVLVGSGLRNIQEMPIPFKIISYFTFQKYCSEIL 180
Qy          576 VVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRTMTNFLILYSFIPALV 635
Db          181 VVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRTMTNFLILYSFIPALV 240
Qy          636 ILGIVVFKIRDHLISR 651
Db          241 ILGIVVFKIRDHLISR 256

```

RESULT 7

US-09-989-981A-4

```

; Sequence 4, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse ABCG8 (mABCG8)
US-09-989-981A-4

```

```

Query Match          21.0%; Score 697; DB 10; Length 672;
Best Local Similarity 29.1%; Pred. No. 1.7e-57;
Matches 195; Conservative 129; Mismatches 263; Indels 84; Gaps 18;

```

```

Qy          15 LQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVR-----PWWD-ITSCR 61
Db          17 LQDASGLQDSL----FSSES DNSLYFTYSGQSNLTLEVRDLTYQVDIASQVPWFQQLAQFK 72

```

Qy	62	QQWTRQI-----LKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGE	112
Db	73	IPWRSHSSQDSCELGIRNLSFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKMKSGQ	131
Qy	113	VYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAI-RRGNPGSFQKKVEAV	171
Db	132	IWINGQPSTPQLVRKCAVHVRQHDQLLPNLTVRETAFIAQMLPRTFSQAQRDKRVEDV	191
Qy	172	MAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQI	231
Db	192	IAELRLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGILILDEPTSGLDSFTAHL	251
Qy	232	VVLLVELARRNRIVVLTIHQPRSELFQLEFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPC	291
Db	252	VTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVOYFTSIGHPC	311
Qy	292	PEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS-----ICHKTLKNIERM	345
Db	312	PRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTS	371
Qy	346	KHLKTLPMVPFKTKDS-----PGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMG	397
Db	372	THTVSLTL---TQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACLMS	427
Qy	398	LFLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVLRVAVSDQESQDGL	457
Db	428	LIIGFLYYGHGAKQL--SFMDTAALLFMIGALIPFNVILDVVSCKHSERSMLYYELEDGL	485
Qy	458	YQKWQMLLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFL-	516
Db	486	YTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELF-----LL--HFLLVWL	537
Qy	517	-----TLVLLGIVQNPNI-VNSVVALLSIAGVLVSGFGLRNIQEMPIPFKIISYFTFQKY	570
Db	538	VFCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW	597
Qy	571	CSEILVVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRTMNFLILY	628
Db	598	CFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMDLNSHPLY	640
Qy	629	SFIPALVILGI	639
Db	641	AIY--LIVIGI	649

## RESULT 8

US-09-989-981A-8

; Sequence 8, Application US/09989981A

; Publication No. US20030049730A1

; GENERAL INFORMATION:

; APPLICANT: Hobbs, Helen H.

; APPLICANT: Shan, Bei

; APPLICANT: Barnes, Robert

; APPLICANT: Tian, Hui

; APPLICANT: Tularik Inc.

; APPLICANT: Board of Regents, The University of Texas System

; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use

```
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human ABCG8 (hABCG8)
US-09-989-981A-8
```

```
Query Match          21.0%; Score 697; DB 10; Length 673;
Best Local Similarity 28.9%; Pred. No. 1.7e-57;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;
```

```
Qy      8 TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQQW 64
      || : ||| | | | : :||| : :| | : ||: : : |
Db     16 TPQDTSGLQDRLFSSSEDNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFQQLAQFKMPW 75

Qy     65 TRQI-----LKDVSLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTF-LGEVYV 115
      | :||| | |||: |:||| | :||| :||| | | | :||:
Db     76 TSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134

Qy    116 NGRALRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAI-RRGNPGSFQKKVEAVMAE 174
      ||: : : | :|| | : || :||| ||| : | : : | : | :|| | :||
Db    135 NGQPSSPQLVRKCVAHVRQHNQLLPLNTLVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAE 194

Qy    175 LSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVL 234
      | | || :|| : | :| ||||| ||| :| :|| :||| ||| ||: :|
Db    195 LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKT 254

Qy    235 LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEPH 294
      | ||: ||:| :||| :||| :||| : :|| | :| | : :| |||| :
Db    255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314

Qy    295 SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHL----- 348
      ||| |||: |||: | :||: ||: | :|| : : : : :||: |
Db    315 SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362

Qy    349 -----KTLPM----VPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRL 390
      | : :| || || : | :|| | | : :
Db    363 AETKDLDEDTCVESSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421

Qy    391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFPPVLR 446
      : :| : : | :|| | ||| : :|| : :|| : :
Db    422 AEACLSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVLIDVISKCYSER 475

Qy    447 AVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARF----- 499
      | : | :||| | | || :|| | | : |
Db    476 AMLYIELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLV 535
```

Qy 500 -----GYFSAALLAPHLIGEFLLTLVLLGIVQNPNIIVNSVVALLSIAGVLVSGGFL 549  
 :||| : | : |:: :|| ||:  
 Db 536 WLTVFCCRIMALAAAALLPTFHMASFFS-----NALYNSFYLAG-----GFM 577  
 Qy 550 RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597  
 |: : || :| :| | |: :| : |:: :  
 Db 578 INLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVS 625

RESULT 9

US-10-090-455-7

; Sequence 7, Application US/10090455  
 ; Publication No. US20030027259A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chen, Hongyun  
 ; APPLICANT: Le Bihan, Stephane  
 ; TITLE OF INVENTION: NOVEL ABCG4 TRANSPORTER AND USES THEREOF  
 ; FILE REFERENCE: 100103.406  
 ; CURRENT APPLICATION NUMBER: US/10/090,455  
 ; CURRENT FILING DATE: 2002-03-01  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 673  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-090-455-7

Query Match 21.0%; Score 697; DB 14; Length 673;  
 Best Local Similarity 28.9%; Pred. No. 1.7e-57;  
 Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

Qy 8 TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQOW 64  
 || : ||| | | : :||| : :| | : ||:: : : |  
 Db 16 TPQDTSGLDRLFSSES DNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFELAQFKMPW 75  
 Qy 65 TRQI-----LKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTF-LGEVYV 115  
 | :::| | ||:: |: ||| |: :||| :|| | | |:::  
 Db 76 TSPSCQNSCELGIQNL SFKVRSGQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134  
 Qy 116 NGRALRREQFQDCFSYVLQSDTLLSSLTVRETLYHTALLAI-RRGNPGSFQKKVEAVMAE 174  
 ||: : : | :| | : || : ||||| : | : : | : | || |::|  
 Db 135 NGQPSSPQLVRKCAVHVRQHNLPLNTLVRETAFIAQMRLPRTFSQAQRDKRVEDVIAE 194  
 Qy 175 LSLSHVADRLIGNYSLGGISTGERRRV SIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVL 234  
 | | || :|| : |:| ||||| ||| :| ::: |||: || |:: :|  
 Db 195 LRLRQCADTRVGNMYVRGLSGGERRRV SIGVQLLWNPGLILDEPTSGLD SFTAHLNLT 254  
 Qy 235 LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEH 294  
 | ||: ||: ||::||| ||::||| : ::: | | : | : :| |||| :  
 Db 255 LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314  
 Qy 295 SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS AICHKTLKNIERMKHL----- 348  
 ||| |||: |||: | :|:|: | : : | : : : :|::: |  
 Db 315 SNPADFVYDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362  
 Qy 349 -----KTLPM----VPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRL 390

```

      |:  :| || ||  :  |:||  |  |:  ::
Db      363 AETKDLDEDTCEVSSVTPLDTNCLPSPTK-MPGAVQQFTTLIRRQISNDFRDLPTLLIHG 421
Qy      391 LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFPLVR 446
      :  :| : : |      |:||  |  ||:      |:  :|: ::  |
Db      422 AEACILMSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVLVDVISKCYSER 475
Qy      447 AVSDQESQDGLYQKWQMLLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVARF----- 499
      |:  | :|||  |  | ||  :|:  ||  |  | :  |
Db      476 AMLYYELEDGLYTTGPYFFAKILGELPEHCAYIIIIYGMPTYWLANLRPGLQPFLHFLLV 535
Qy      500 -----GYFSAALLAPHLIGEFTLVLLGIVQNPNIIVNSVALLSIAGVLVGSGL 549
      :|||  :  | :      |::  :||  ||:
Db      536 WLNVFCCRIMALAAAALLPTFHMASFFS-----NALYNSFYLAG-----GFM 577
Qy      550 RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597
      |:  :  || :| :| | | :|  :  |::  :
Db      578 INLSSLWTPPAWISKVSVFLRWCFEGLMKIQFSRRTYKMPLGNLTIAS 625

```

# RESULT 10

US-09-961-086-1

```

; Sequence 1, Application US/09961086
; Publication No. US20030036645A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF MARYLAND, BALTIMORE
; APPLICANT: ROSS, Douglas D.
; APPLICANT: DOYLE, L. Austin
; APPLICANT: ABRUZZO, Lynne
; TITLE OF INVENTION: BREAST CANCER RESISTANCE PROTEIN (BCRP) AND THE DNA
; TITLE OF INVENTION: WHICH ENCODES IT
; FILE REFERENCE: EP19376-019
; CURRENT APPLICATION NUMBER: US/09/961,086
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/073,763
; PRIOR FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US99/02577
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-086-1

```

```

Query Match          20.5%;  Score 682.5;  DB 10;  Length 655;
Best Local Similarity 29.2%;  Pred. No. 4e-56;
Matches 182;  Conservative 138;  Mismatches 249;  Indels 55;  Gaps 18;

```

```

Qy      21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77
      || :  | |||  :  :| : : :: :||:      ||:  :|| ::  ::
Db      13 SQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK 72
Qy      78 SGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFDQCFSYVLQSDT 137
      | :  ||| :| ||::||| ::  |  :|  |:| :|| |  | :  ||:| |
Db      73 PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLING-APRPANFKCNSGYVVQDDV 129

```



; SEQ ID NO 13  
; LENGTH: 655  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: ABCG2 482Tmutant sequence  
US-10-405-806-13

Query Match 20.5%; Score 682.5; DB 15; Length 655;  
Best Local Similarity 29.2%; Pred. No. 4e-56;  
Matches 182; Conservative 138; Mismatches 249; Indels 55; Gaps 18;

Qy	21	SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE	77
		:         : :   : : : :     :   :   : : :	
Db	13	SQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK	72
Qy	78	SGQIMCILGSSSGSKTTLDDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT	137
		:       :       :     :   :   :       :	
Db	73	PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLING-APRPANFKCNSGYVQDDV	129
Qy	138	LLSSLTVRETLHYTALLAIRRGNGP-SFQKKVEAVMAELSLSHVADRLLIGNYSLGGISTG	196
		: : :           : :   : : : :   :           :   :   :	
Db	130	VMGTLTVRENLOFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIRGVSGG	189
Qy	197	ERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVLTIHQPRSEL	256
		:     :   : :               : :   : : :   : :         :	
Db	190	ERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMKQGRTIIFSIIHQPRYSI	249
Qy	257	FQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQ----SK	312
		:     : :   :   :     :       :     :     : : :   : : : :	
Db	250	FKLFDSLTLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR	309
Qy	313	ERE-----IETSKR----VQMIESAYKKSACHKT-----LKNIERMKHLKTLPMVPF	356
		: : :       : : : :   : :   : :   :   : : : : : : :	
Db	310	EEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY	369
Qy	357	KTKDSPGVFESKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFVLRVRSNVLKG	414
		: : :   : :   :   : : : : : : : : :   :   :	
Db	370	TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDST----	421
Qy	415	AIQDRVGLLYQFVGATPYTGMLNAVNLFVPVLRVSDQESQDGLYQKWQMMLAYAL-HVLP	473
		:   : :   : : :       : :   :   :   :   :   :	
Db	422	GIQNRAGVLF-FLT TNQCFSSVSAVELFVVEKKLFIHEYISGYRYRVSSYFLGKLLSDLLP	480
Qy	474	FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNI VNSV	533
		: : : :   : :   :     :   : : : : : : : :   : : : :	
Db	481	MTMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVVSV	537
Qy	534	VALLSIAGV--LVGSGFLRNIQEMPIPFKII SYFTFQKYCSEILVNEFYGLNFTCGSSN	591
		:     : :       : : : :   : :   :	
Db	538	TLLMTICFVFMMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNFPCPG---	594
Qy	592	VSVTTNPMCAFTQGIQFIEKTCPG	615
		: :       :	
Db	595	LNATGNNPCNYA-----TCTG	610

RESULT 12  
 US-09-981-353-35  
 ; Sequence 35, Application US/09981353  
 ; Patent No. US20020160382A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lasek, Amy W.  
 ; APPLICANT: Jones, David A.  
 ; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER  
 ; FILE REFERENCE: PA-0038 US  
 ; CURRENT APPLICATION NUMBER: US/09/981,353  
 ; CURRENT FILING DATE: 2001-10-11  
 ; NUMBER OF SEQ ID NOS: 194  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 35  
 ; LENGTH: 655  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; OTHER INFORMATION: Incyte ID No. US20020160382A1 5517972CD1  
 US-09-981-353-35

Query Match 20.5%; Score 680.5; DB 9; Length 655;  
 Best Local Similarity 29.2%; Pred. No. 6.1e-56;  
 Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

Qy	21	SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSPLYVE	77
		:         : :   : : : :     : :     : : : :	
Db	13	SQGNTNGFPATASNDLKAFTGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK	72
Qy	78	SGQIMCILGSSSGSGKTTLLDAMSGR LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT	137
		:       :       :     : :   :   :       :   :	
Db	73	PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDV LING-APRPANFKCNSGYVVQDDV	129
Qy	138	LLSSLTVRETLHYTALLAIRRGNPG-SFQKKVEAVMAELSLSHVADRLIGNYSLGGISTG	196
		: : :           : :   : : : :   :           :   :   :	
Db	130	VMGTLTVREN LQFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIRGVSGG	189
Qy	197	ERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSEL	256
		:     :   :               : :     : : :   : : :	
Db	190	ERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIFSIHQPRYSI	249
Qy	257	FQLFDKIAILSFGELIFCGTPAEMLDFFND CGYPCPEHSNPFDFYMDLTSVDTQ----SK	312
		:       : :   :     :     :     :     :     :   : : :	
Db	250	FKLFDSLTL LASGR LMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR	309
Qy	313	ERE-----IETSKR----VQMI ESAYKKS AICHKT-----LKNIERMKHLKTLPMVPF	356
		: : :         : : :   : :   : :   : :   : :   : : :	
Db	310	EEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY	369
Qy	357	KTKDSPGVFSKLGVL LRRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFVLRVRSNV LKG	414
		: : :   : :     : : : : :   : :   : :   : :	
Db	370	TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKNDST----	421
Qy	415	AIQDRVGLLYQFVGATPYTGMLNAVNLF PVLRAVSDQESQDGLYQKWQMLLAYAL-HVLP	473
		:   :   : : :       : :   :   :   :   :	
Db	422	GIQNRAGVLF-FLT TNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP	480



Qy 474 FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSV 533  
 :: ::||: : |: ||| |: | |: :: : : | | ::|:  
 Db 481 MRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVVSV 537  
 Qy 534 VALLSIAGV--LVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSN 591  
 |::| | : : || | |: : : ||: :| | || | || |  
 Db 538 TLLMTICFVFMMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNFPCPG--- 594  
 Qy 592 VSVTTNPMCAFTQGIQFIEKTCPG 615  
 :: | | | : || |  
 Db 595 LNATGNNPCNYA-----TCTG 610

RESULT 13

US-10-120-687-61

; Sequence 61, Application US/10120687  
 ; Publication No. US20030082155A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Massachusetts General Hospital  
 ; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating Diabetes  
 ; TITLE OF INVENTION: Mellitus  
 ; FILE REFERENCE: 3284/1235B  
 ; CURRENT APPLICATION NUMBER: US/10/120,687  
 ; CURRENT FILING DATE: 2002-04-11  
 ; PRIOR APPLICATION NUMBER: US60/169082  
 ; PRIOR FILING DATE: 1999-12-06  
 ; PRIOR APPLICATION NUMBER: US 09/963,875  
 ; PRIOR FILING DATE: 2001-09-25  
 ; PRIOR APPLICATION NUMBER: US 60/215109  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: US 60/238880  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: US 09/731261  
 ; PRIOR FILING DATE: 2000-12-06  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 61  
 ; LENGTH: 655  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-120-687-61

Query Match 20.5%; Score 680.5; DB 14; Length 655;  
 Best Local Similarity 29.2%; Pred. No. 6.1e-56;  
 Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

Qy 21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77  
 || : | |||| : : | : : :: ::||: ||: ::|| :: :  
 Db 13 SQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK 72  
 Qy 78 SGQIMCILGSSSGSKTTLLDAMSGR LGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT 137  
 | : ||| :| ||::||| :: | :| |:| :|| | | : ||:| |  
 Db 73 PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLING-APRPANFKCNSGYVVQDDV 129  
 Qy 138 LLSSLTVRETLHYTALLAIRRGNPG-SFQKKVEAVMAELSLSHVADRLIGNYSLGGISTG 196

```

Db      130 VMGTLTVRENLQFSAALRLATMTNHEKNERINRVIQELGLDKVADSKVGTQFIRGVSGG 189
Qy      197 ERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSEL 256
      ||:|| || :||: || :| ||||| ||| :||| :||| :||| :|
Db      190 ERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMKQGRTIIFSIIHQPRYSI 249
Qy      257 FQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQ----SK 312
      |:||| :|:| | |:| | | :| || | :||| ||:|:| :| :| :|
Db      250 FKLFDLTLASGRMLMFHGPQAEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR 309
Qy      313 ERE-----IETSKR----VQMIESAYKKSACHKT-----LKNIERMKHLKTLPMVPF 356
      | : || ||: :| :| :| :| :| :| :| :| :| :| :|
Db      310 EEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY 369
Qy      357 KTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFVLRVRSNVLKG 414
      | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db      370 TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIVTVVLGLVIGAIYFGLKNDST---- 421
Qy      415 AIQDRVGLLYQFVGATPYTGMLNAVNLFPPVLRAVSDQESQDGLYQKWQMLAYAL-HVLP 473
      ||:|| |:|: |: :||| || | :| :| :| :| :| :| :| :|
Db      422 GIQNRAGVLF-FLTNNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
Qy      474 FSVVATMIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNIIVNSV 533
      :| :||: :| :||| |: | :| :| :| :| :| :| :| :|
Db      481 MRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVVSVA 537
Qy      534 VALLSIAGV--LVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSN 591
      |:| | :| || | :| :| :| :| :| :| :| :| :| :|
Db      538 TLLMTICFVFMIFSGLLVNLTIASWLSWLQYFSIPRYGFTALQHNEFLGQNFCEPG--- 594
Qy      592 VSVTTNPMCAFTQGIQFIEKTCPG 615
      :| | | :| :| :| :| :| :| :| :| :| :| :| :|
Db      595 LNATGNNPCNYA-----TCTG 610

```

RESULT 14

US-10-405-806-2

; Sequence 2, Application US/10405806

; Publication No. US20030232362A1

; GENERAL INFORMATION:

; APPLICANT: KOMATANI, HIDEYA

; APPLICANT: HARA, YOSHIKAZU

; APPLICANT: KOTANI, HIDEHITO

; APPLICANT: NAKAGAWA, RINAKO

; TITLE OF INVENTION: DRUG RESISTANT GENE AND USE THEREOF

; FILE REFERENCE: 234985US0CONT

; CURRENT APPLICATION NUMBER: US/10/405,806

; CURRENT FILING DATE: 2003-04-03

; PRIOR APPLICATION NUMBER: PCT/JP01/08112

; PRIOR FILING DATE: 2001-09-18

; PRIOR APPLICATION NUMBER: JP2000-303441

; PRIOR FILING DATE: 2000-10-03

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 2

; LENGTH: 655

; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-405-806-2

Query Match 20.5%; Score 680.5; DB 15; Length 655;  
Best Local Similarity 29.2%; Pred. No. 6.1e-56;  
Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

```
Qy      21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77
      || : | ||| : : | : : : : || : || : : || : :
Db      13 SQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK 72

Qy      78 SGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT 137
      | : ||| : | || : || : : | : | : | : | : | : || : |
Db      73 PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLING-APRPANFKCNSGYVVDQDV 129

Qy     138 LLSSLTVRETLHYTALLAIRGNPG-SFQKKVEAVMAELSLSHVADRLIGNYSLGGISTG 196
      : : ||||| | : || : : : : : : || : || : | : | |
Db     130 VMGTTLTVRENLQFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIRGVSGG 189

Qy     197 ERRRVSIAAQLLDQPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSEL 256
      || : || : || : || : ||||| || : || : : : | : : |||| :
Db     190 ERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMSKQGRTIIFSIGHQPRYSI 249

Qy     257 FQLFDKIAILSFGEILFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQ----SK 312
      | : ||| : : | : | : | : | : || : || : || : : | : :
Db     250 FKLFDSLTLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR 309

Qy     313 ERE-----IETSKR----VQMIESAYKKSACHKT-----LKNIERMKHLKTLPMVPF 356
      | : || || : : : | : : | : | : | : | : : :
Db     310 EEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY 369

Qy     357 KTKDSPGVFVSKLGVLLRRVTRNLVRNKLAVITRLLQNLMGLFL--LFFVLRVRSNVLKG 414
      | : : : | : || : | | : : : || : : || : | : |
Db     370 TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKNDST---- 421

Qy     415 AIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYAL-HVLP 473
      || : | : | : : || || | : : | | : | | : ||
Db     422 GIQNRAGVLF-FLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480

Qy     474 FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNI VNSV 533
      : : : || : : | : || | : | : : : : : : | : : | :
Db     481 MRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVVSVA 537

Qy     534 VALLSIAGV--LVGSGFLRNIQEMPIPFKII SYFTFQKYCSEILVNEFYGLNFTCGSSN 591
      | : | | : || | | : : : || : : | || | || |
Db     538 TLLMTICFVFMIFSGLLVNLTITIASWLSWLQYFSIPRYGFTALQHNEFLGQNFPCPG--- 594

Qy     592 VSVTTNPMCAFTQGIQFIEKTCPG 615
      : : | | | : || |
Db     595 LNATGNNPCNYA-----TCTG 610
```

RESULT 15  
US-09-866-866A-10  
; Sequence 10, Application US/09866866A  
; Patent No. US20020102244A1

```
; GENERAL INFORMATION:
; APPLICANT: Sorrentino, Brian
; APPLICANT: Schuetz, John
; TITLE OF INVENTION: A Method of Identifying and/or Isolating Stem Cells
; FILE REFERENCE: 1340-1-021CIP2
; CURRENT APPLICATION NUMBER: US/09/866,866A
; CURRENT FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: 09/584,586
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: PCT/US99/11825
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: 60/086,988
; PRIOR FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-866-866A-10
```

```
Query Match          20.3%; Score 674.5; DB 9; Length 655;
Best Local Similarity 29.0%; Pred. No. 2.3e-55;
Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;
```

```
Qy      21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77
      || : | ||| : : | : : :: :||: ||: :|| :: :
Db      13 SQGNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKCLKSGFLPCRKPVEKEILSNINGIMK 72

Qy      78 SGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT 137
      | : ||| :| ||::||| :: | :| | :| :| | | : ||:| |
Db      73 PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLING-APRPANFKCNSGYVVQDDV 129

Qy     138 LLSSLTVRETLHYTALLAIRRGNGP-SFQKKVEAVMAELSLSHVADRLLIGNYSLGGISTG 196
      :: :||| | :| | : :: | : || | || | : | : | |
Db     130 VMGTLTVRENLOQFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQFIRGVSGG 189

Qy     197 ERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVLLVELARRNRIVLTIHQPRSEL 256
      ||:| || :|: || :: ||||| ||| ::|| :::: | :: :||| | :
Db     190 ERKRTSIGMELITDPSILSLDEPTTGLDSSTANAVLLLLKRMKQGRTIIFSIIHQPRYSI 249

Qy     257 FQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQ----SK 312
      |:| | : :| :| | | :| || | :|| ||::| : : | : ::
Db     250 FKLFDLTLASGRMLFMHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR 309

Qy     313 ERE-----IETSKR----VQMIESAYKKSACHKT-----LKNIERMKHLKTLPMVPF 356
      | : || ||: :: : | | : :| | | : | : | : : :
Db     310 EEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY 369

Qy     357 KTKDSPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFVLRVRSNVLKG 414
      | : :| : :| :||: | | | :: ::|| : :| | : |
Db     370 TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKNDST---- 421

Qy     415 AIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYAL-HVLP 473
      ||:| |:| :| :||| || | : : | | : | | :||
Db     422 GIQNRAGVLF-FLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480
```

Qy	474	FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSV	533
		:: ::  : :  :      :   : :: : :     :: :	
Db	481	MRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVVSV	537
Qy	534	VALLSIAGV--LVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSN	591
		::    ::     : : :   : :	
Db	538	TLLMTICFVFMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNF	594
		CPG---	
Qy	592	VSVTTNPMCAFTQGIQFIEKTC	615
		::       :	
Db	595	LNATGNNPCNYA-----TCTG	610

Search completed: February 27, 2004, 07:34:06  
 Job time : 30.2557 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 36.1394 Seconds  
(without alignments)  
5683.620 Million cell updates/sec

Title: US-09-989-981A-6  
Perfect score: 3326  
Sequence: 1 MGDLSSLTPGGSMGLQVNRG.....PALVILGIVVFKIRDHLISR 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

%  
Result Query  
No. Score Match Length DB ID Description  
-----

1	2726.5	82.0	652	11	Q7TSR8	Q7tsr8 mus musculu
2	704	21.2	673	11	Q8R543	Q8r543 mus musculu
3	701	21.1	672	11	Q7TSR7	Q7tsr7 mus musculu
4	697	21.0	672	11	Q7TSR6	Q7tsr6 mus musculu
5	691	20.8	672	11	Q8CIQ5	Q8ciq5 rattus norv
6	680.5	20.5	655	4	Q96TA8	Q96ta8 homo sapien
7	679.5	20.4	655	4	Q8IX16	Q8ix16 homo sapien
8	672.5	20.2	655	4	Q96LD6	Q96ld6 homo sapien
9	668.5	20.1	656	6	Q8MIB3	Q8mib3 sus scrofa
10	663	19.9	657	11	Q7TMS5	Q7tms5 mus musculu
11	662	19.9	801	5	Q8T691	Q8t691 dictyosteli
12	660	19.8	657	11	Q9R004	Q9r004 mus musculu
13	649.5	19.5	725	10	Q9M3D6	Q9m3d6 arabidopsis
14	639.5	19.2	643	5	Q7YYX5	Q7yyx5 cryptospori
15	628	18.9	691	10	Q8RWI9	Q8rwi9 arabidopsis
16	626.5	18.8	657	11	Q80XF3	Q80xf3 rattus norv
17	623	18.7	679	5	Q8IS30	Q8is30 bactrocera
18	622.5	18.7	657	11	Q80W57	Q80w57 rattus norv
19	622	18.7	668	10	Q9ARU4	Q9aru4 oryza sativ
20	620.5	18.7	657	11	Q80ST1	Q80st1 rattus norv
21	620	18.6	692	10	Q7XUM2	Q7xum2 oryza sativ
22	618.5	18.6	672	10	Q9LI82	Q9li82 arabidopsis
23	617	18.6	727	10	Q9FNB5	Q9fnb5 arabidopsis
24	615.5	18.5	723	10	Q8LNT5	Q8lnt5 oryza sativ
25	615	18.5	692	5	P91892	P91892 aedes aegyp
26	614.5	18.5	703	10	Q8RXN0	Q8rxn0 arabidopsis
27	614	18.5	594	10	Q9LJC3	Q9ljc3 arabidopsis
28	614	18.5	720	10	Q9M2V7	Q9m2v7 arabidopsis
29	610.5	18.4	725	10	Q9ZU35	Q9zu35 arabidopsis
30	610.5	18.4	725	10	Q9ASR9	Q9asr9 arabidopsis
31	610	18.3	679	5	Q9BH97	Q9bh97 ceratitidis c
32	608	18.3	708	10	Q9M2V5	Q9m2v5 arabidopsis
33	602.5	18.1	654	10	Q9LIW2	Q9liw2 oryza sativ
34	600.5	18.1	670	5	O77423	O77423 bactrocera
35	600	18.0	604	5	Q8MRJ2	Q8mrj2 drosophila
36	600	18.0	787	10	Q8H8V7	Q8h8v7 oryza sativ
37	597	17.9	590	10	Q9MAH4	Q9mah4 arabidopsis
38	595.5	17.9	658	5	O16574	O16574 caenorhabdi
39	595.5	17.9	687	5	Q94960	Q94960 drosophila
40	595.5	17.9	785	4	Q96L76	Q96l76 homo sapien
41	592	17.8	610	5	P90746	P90746 caenorhabdi
42	591.5	17.8	740	10	O80946	O80946 arabidopsis
43	589.5	17.7	646	10	Q9C6R7	Q9c6r7 arabidopsis
44	588.5	17.7	646	11	Q8K4E1	Q8k4e1 mus musculu
45	588.5	17.7	648	10	Q9C6W5	Q9c6w5 arabidopsis

#### ALIGNMENTS

#### RESULT 1

Q7TSR8

ID Q7TSR8 PRELIMINARY; PRT; 652 AA.

AC Q7TSR8;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE ATP-binding cassette sub-family G member 5.  
 GN ABCG5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=I/LnJ; TISSUE=Liver;  
 RA Wittenburg H., Lyons M.A., Li R., Churchill G.A., Carey M.C.,  
 RA Paigen B.;  
 RT "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone  
 RT Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred  
 RT Mice.";  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY195872; AA045093.1; -.  
 KW ATP-binding.  
 SQ SEQUENCE 652 AA; 73236 MW; 0125FB617DE296B9 CRC64;

Query Match 82.0%; Score 2726.5; DB 11; Length 652;  
 Best Local Similarity 79.4%; Pred. No. 5.5e-193;  
 Matches 518; Conservative 68; Mismatches 65; Indels 1; Gaps 1;

Qy	1	MGDLSSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS	59
		:   :   :  :               :       :	
Db	1	MGELPFLSPEGARGPHINRGSLSLEQGSVTGTPEARHSLGVLHVSYSVSNRVGPWWNIKS	60
Qy	60	CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA	119
		: :           :             :       : :	
Db	61	CQQKWDRQILKDVSLYIESGQIMCILGSSGSGKTTLLDAISGRRLRCTGTLEGDFVNGCE	120
Qy	120	LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGNPGSFQKKVEAVMAELSLSH	179
		:                       :    : :	
Db	121	LRRDQFQDCFSYVLQSDVFLSSLTVRETLRYTAMLALCRSSADFYNNKKVEAVMTELSLSH	180
Qy	180	VADRLIGNYSLGGISTGERRRVSIQAQLLQDPKVMFLDEPTTGLDCMTANQIVVLLVELA	239
		:   :   :   :   :               :             :	
Db	181	VADQVIGSYNFGGISSGERRRVSIQAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA	240
Qy	240	RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF	299
		:   :   :               :   :   :       :   :	
Db	241	RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNDCGYPCPEHSNPF	300
Qy	300	FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK	359
		:   :         :   :   :       :   :   :   :	
Db	301	FYMDLTSVDTQSREREIETYKRVQMLESFAKESDIYHKILENIERARYLKTLPVPFKTK	360
Qy	360	DSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR	419
		:           :   :       :   :   :   :   :   :   :	
Db	361	DPPGMFGKLGVLRLRRVTRNLVRNKLAVIMRLVQNLIMGLFLIFVLRVQNTLKGAVQDR	420
Qy	420	VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVAT	479
		:   :         :   :   :       :   :	
Db	421	VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSIAT	480
Qy	480	MIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVALLSI	539
		:         :   :         :   :   :   :   :   :   :	









```

Db      486 YTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTNLRPVPELF-----LL--HLLLVLV 537
Qy      517 -----TLVLLGIVQNPNI-VNSVVALLSIAGVLVGSGFLRNIQEMPIPFKIISYFTFQKY 570
          | : |      | : |      : : | : : |      | : | :
Db      538 VFCCRTMALAASAMLPTFHMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW 597
Qy      571 CSEILVVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATS RFTMNFILILY 628
          | : : | : | : | : : |      | : | |
Db      598 CFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMDLNSHPLY 640
Qy      629 SFIPALVILGI 639
          : | : : |
Db      641 AIY--LIVIGI 649

```

## 07TSR6

```

ID      Q7TSR6          PRELIMINARY;          PRT;      672 AA.
AC      Q7TSR6;
DT      01-OCT-2003 (TrEMBLrel. 25, Created)
DT      01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      ATP-binding cassette sub-family G member 8.
GN      ABCG8.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=PERA/Ei; TISSUE=Liver;
RA      Wittenburg H., Lyons M.A., Li R., Churchill G.A., Carey M.C.,
RA      Paigen B.;
RT      "Primary Roles of FXR and ABCG5/ABCG8 in Cholesterol Gallstone
RT      Susceptibility: Evidence from a Cross of PERA/Ei and I/Ln Inbred
RT      Mice.";
RL      Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR      EMBL; AY196216; AAO45096.1; -.
KW      ATP-binding.
SQ      SEQUENCE      672 AA;  75867 MW;  CAB720502EA8FE21 CRC64;

```

Query Match 21.0%; Score 697; DB 11; Length 672;  
Best Local Similarity 29.1%; Pred. No. 5.9e-43;  
Matches 195; Conservative 129; Mismatches 263; Indels 84; Gaps 18;

Qy	15	LQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVR-----PWWD-ITSCR	61
		: : :   :    :: : :	
Db	17	LQDASGLQDSL----FSSESDNSLYFTYSGQNTLEVRDLTYQVDIASQVPWFQLAQFK	72
Qy	62	QQWTRQI-----LKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTF-LGE	112
		::::      ::  :      : :    :        : :	
Db	73	IPWRSHSSQDSCELGIRNLSFKVRSQMLAIIGSSGCGRASLLDVITGR-GHGGKMKSGQ	131
Qy	113	VYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAI-RRGNPGSFQKKVEAV	171
		:::  : :   : :       :       :   : :   :   :	
Db	132	IWINGQPSTPOLVRKCVAHVROHDQLLPNLTVRETLAFIAQMRLPRTFSAOORDKRVEDV	191

Qy 172 MAELSLSHVADRLIGNYSLGGISTGERRRVSIQAQLLQDPKVMFLFDEPTTGLDCMTANQI 231  
 :||| | : :|| : |:| ||||| ||| :| :: |||:| | ||: :  
 Db 192 IAE LR LRQC ANTRV GNTYVRGVSGGERRRVSIGVQLLWNP GILILDEPTSGLD SFTA HNL 251  
 Qy 232 VVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFND CGYPC 291  
 | | ||: ||:|:::| |||:|:| | : :: | |: | :|: | |:|  
 Db 252 VTTLSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQQMVQYFTSIGHPC 311  
 Qy 292 PEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS A-----ICHKTLKNIERM 345  
 | :||| |||:| |||:| :||| | : : | : : : : : | :  
 Db 312 PRYSNPADFYVDLTSIDRRSKEREVATVEKAQSLAALFLEKVQGFDDFLWKAEAKELNTS 371  
 Qy 346 KHLKTLPMVPFKTKDS-----PGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMG 397  
 | :| : |:| : ||: : |:| | |: : : : :|  
 Db 372 THTVSLTL---TQDTCGTAVELPGMIEQFSTLIRRQISNDFRDLPTLLIHGSEACIMS 427  
 Qy 398 LFLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVLRVAVSDQESQDGL 457  
 | : | : | : | ||: |: :|: | :|: | :|  
 Db 428 LIIGFLYYGHGAKQL--SFMDTAALLFMIGALIPFNVILDVVS KCHSERSMLY ELEDGL 485  
 Qy 458 YQKWQMLLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFL- 516  
 | | | | :|: || | | | | | : :|  
 Db 486 YTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVPELF-----LL--HFLLVWL V 537  
 Qy 517 -----TLVLLGIVQNPNI-VNSVVALLSIAGVLVSGFLRNIQEMPPIPKIISYFTQKY 570  
 |: | | :| : :|: |: : | | :| : :  
 Db 538 VFCCRTMALAASAMLP TFMSSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLSFLRW 597  
 Qy 571 CSEILVVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATS RFTMNF LILY 628  
 | |: :| | :| |: |: : | | :| ||  
 Db 598 CFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMDLNSHPLY 640  
 Qy 629 SFIPALVILGI 639  
 : |::||  
 Db 641 AIY--LIVIGI 649

# RESULT 5

## Q8CIQ5

ID Q8CIQ5 PRELIMINARY; PRT; 672 AA.  
 AC Q8CIQ5;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Sterolin 2.  
 GN ABCG8.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Yu H., Lu K., Lee M., Pandit B., Patel s.B.;  
 RT "The rat Abcg5 and Abcg8: characterization, chromosomal assignment and  
 RT genetic variation in sitosterolemic rats."  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY145899; AAN64276.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 SQ SEQUENCE 672 AA; 75906 MW; 2FE0846E71BD9D47 CRC64;

Query Match 20.8%; Score 691; DB 11; Length 672;  
 Best Local Similarity 28.3%; Pred. No. 1.6e-42;  
 Matches 189; Conservative 126; Mismatches 264; Indels 88; Gaps 15;

Qy	23	SSLEGAPATAPEPHSLGILHASYSVSHRVR-----PW-----WDITSC	60
		: : : :     : :   :	
Db	21	SSLQDSVFSSES DNSLYFTYSGQSNLTLEVRDLTYQVDMASQVPWF EQLAQFKLPWRSRGS	80
Qy	61	RQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
		:       : : :         :   :   :   :   :   :   :   :   :   :	
Db	81	QDSWDLGI-RNLSFKVRSGQMLAIIGSAGCGRATLLDVITGRDHGGKMKSGQIWINGQPS	139
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLLHYTALLAIRGNPGSF-----QKKVEAVMAEL	175
		:     : :         :           :   : :   :     :	
Db	140	TPQLIQKCVAHVRQQDQLLPNLTVRETTLFIAQMRL----PKTFSQAQRDKRVEDVIAEL	195
Qy	176	SLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLEFDEPTTGLDCMTANQIVVLL	235
		: : :   :                   :   : :       :	
Db	196	RLRQCANTRVGNTYVRGVSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVRTL	255
Qy	236	VELARRNRIVVLTIIHQPRSELFQFLDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHS	295
		:   : : :         :       : : :     :     :         :	
Db	256	SRLAKGNRLVLI SLHQPRSDIFRLFDLVLLMTSGTPIYLGVAQH MVQYFTSIGYPCPRYS	315
Qy	296	NPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKS-----ICHKTLKNIERMKHLK	349
		:       :   :   : : : : : : : : : :   : :	
Db	316	NPADFYVDLTSIDRRSKEQEVATMEKARLLAALFLEKVQGFDDFLWKAEAKSLD----TG	371
Qy	350	TLPMVPFKTKDS-----PGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLMGLFLL	401
		:   :   :   : :   :       : : : :   : :   :	
Db	372	TYAVSQTLTQDTNCGTAAELPGMIQQFTTLIRRQISNDFRDLPTLFIHGAEACLM SLIIG	431
Qy	402	FFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVAVSDQESQDGLYQKW	461
		:   :   :   : :   : : :   : :   :	
Db	432	FLYYGHADKPL--SFMDMAALLFMIGALIPFNVILDVVS KCHSERSLLYYELEDGLYTAG	489
Qy	462	QMMLAYALHVL PFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEF-----	516
		: :                 :	
Db	490	PYFFAKVLGELPEHCAYVIIYGMPIYWLTLNLRP-----GPELFLHFMLLWLWLVVFCC	541
Qy	517	-TLVLLGIVQNPNI-VNSVVALLSIAGVLVSGSGLRN IQEMPIPKIISYFTFQKYCSEI	574
		:     : :   : :   : :     :   :	
Db	542	RTMALAASAMLPTFHMSSFCCNALYNSFYLTAGFMINLNLWIVPAWISKMSFLRWCFSG	601
Qy	575	LVVNEFYGLNFTCGSSNSVSTTNPMCAFTQGIQFIEKTCPG--ATSRFTMNFLLILYSFIP	632

```

      |:  :| |  :|  |::  :      ||  :  :|  ||:
Db      602 LMQIQFNHGIYTTQIGNLTFSV-----PGDAMVTAMDNLNSHPLYAIY- 643

Qy      633 ALVILGI 639
      |::|||
Db      644 -LIVIGI 649

```

RESULT 6

Q96TA8

```

ID   Q96TA8          PRELIMINARY;          PRT;   655 AA.
AC   Q96TA8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   ATP-binding cassette superfamily G (White) member 2 (Hypothetical
DE   protein).
GN   ABCG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21201983; PubMed=11306452;
RA   Komatani H., Kotani H., Hara Y., Nakagawa R., Matsumoto M.,
RA   Arakawa H., Nishimura S.;
RT   "Identification of breast cancer resistant protein/mitoxantrone
RT   resistance/placenta-specific, ATP-binding cassette transporter as a
RT   transporter of NB-506 and J-107088, topoisomerase I inhibitors with an
RT   indolocarbazole structure.";
RL   Cancer Res. 61:2827-2832(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreatic carcinoma;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB051855; BAB46933.1; -.
DR   EMBL; BC021281; AAH21281.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00005; ABC_tran; 1.
DR   ProDom; PD000006; ABC_transporter; 1.
DR   PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.
KW   Hypothetical protein; ATP-binding.
SQ   SEQUENCE   655 AA;  72314 MW;  A8AF66B96034C5A8 CRC64;

```

```

Query Match          20.5%;  Score 680.5;  DB 4;  Length 655;
Best Local Similarity 29.2%;  Pred. No. 9.4e-42;
Matches 182;  Conservative 137;  Mismatches 250;  Indels 55;  Gaps 18;

```

Qy 21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77





RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yoshikawa M., Yabuuchi H., Ikegami Y., Ishikawa T.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AF463519; AA014617.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 655 AA; 72314 MW; A8AF60B591D4C5A8 CRC64;

Query Match 20.4%; Score 679.5; DB 4; Length 655;  
 Best Local Similarity 29.2%; Pred. No. 1.1e-41;  
 Matches 182; Conservative 137; Mismatches 250; Indels 55; Gaps 18;

Qy 21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77  
 || : | ||| : : | : : :: :|| : || : :|| :: :  
 Db 13 SQGNTNGFPATASNDLKAFTGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK 72  
  
 Qy 78 SGQIMCILGSSSGSKTTLDDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT 137  
 | : ||| :| ||: ||| :: | :| | :| :|| | | :| ||: | |  
 Db 73 PG-LNAILGPTGGGKSSLLDVLAARKDPSG-LSGDVLING-APRPANFKCNSGYVQDDV 129  
  
 Qy 138 LLSSLTVRETLHYTALLAIRGNPG-SFQKKVEAVMAELSLSHVADRILIGNYSLGGISTG 196  
 :: :||| | :| | : :: :| : || | ||| :| : | :| |  
 Db 130 VMGTLTVRENLFSAALRLATTMTNHEKNERINRVIQELGLDKVADSKVGTQFIRGVSGG 189  
  
 Qy 197 ERRRVSIAAQLLQDPKVMFLDEPTTGLDCMTANQIVLLVELARRNRIVLTIHQPRSEL 256  
 ||: | || :| : || :: ||||| ||| ::|| :::: | :: :||| :  
 Db 190 ERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMKQGRITIFSIIHQPRYSI 249  
  
 Qy 257 FQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQ----SK 312  
 | :||| : :| : | :| | | :| || | :|| ||:| : :| : :  
 Db 250 FKLFDSLTLASGRMLFMHGPQAEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR 309  
  
 Qy 313 ERE-----IETSKR----VQMIESAYKKSACHKT-----LKNIERMKHLKTLPMVPF 356  
 | : || ||| : : : | | :| :| :| :| :| :| :| : : :  
 Db 310 EEDFKATEIIEPSKQDKPLIEKLAIEYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY 369  
  
 Qy 357 KTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFVLRVRSNVLKG 414  
 | :| : :| :|| : | | :| :| :| :| :| :| :| :| :| :|  
 Db 370 TT----SFCHQLRWVSKRSFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKNDST---- 421  
  
 Qy 415 AIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLAYAL-HVLP 473  
 ||: | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Db 422 GIQNRAGVLF-FLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSSYFLGKLLSDLLP 480

Qy 474 FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSV 533  
 :: ::||: : |: ||| |: | |: :: : : | | ::|:  
 Db 481 MRMLPSIIFTTCIVYFMLGLKPKADAFFVMMFTLM--MVAYSASSMALAIAAGQSVVSV 537  
 Qy 534 VALLSIAGV--LVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSN 591  
 |::| | : : || | |: : : ||: :| | || | || |  
 Db 538 TLLMTICFVEMMIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNFPCPG--- 594  
 Qy 592 VSVTTNPMCAFTQGIQFIEKTCPG 615  
 :: | | | : || |  
 Db 595 LNATGNNPCNYA-----TCTG 610

# RESULT 8

Q96LD6

ID Q96LD6 PRELIMINARY; PRT; 655 AA.  
 AC Q96LD6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ABC transporter ABCG2.  
 GN ABCG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Schuetz J.D., Wall A.M., Sampath J., Sorrentino B., Du G.;  
 RT "The Human ABC Transporter, ABCG2, Transports Hoechst 33342 and  
 RT Requires an Intact Walker A Motif."  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AY017168; AAG52982.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 655 AA; 72288 MW; B3B5DC02C095C4A8 CRC64;

Query Match 20.2%; Score 672.5; DB 4; Length 655;  
 Best Local Similarity 29.0%; Pred. No. 3.7e-41;  
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

Qy 21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77  
 || : | |||| : : | : : :: :||: ||: ::|| :: : :  
 Db 13 SQGNTNGFPATASNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK 72



RA Eisenblaetter T., Galla H.J.;  
 RT "A new multidrug resistance protein at the blood-brain barrier."  
 RL Biochem. Biophys. Res. Commun. 293:1273-1278(2002).  
 DR EMBL; AJ420927; CAD12785.1; -.  
 DR PIR; JC7860; JC7860.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
 DR GO; GO:0000166; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR006162; Ppantne\_S.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
 KW ATP-binding.  
 SQ SEQUENCE 656 AA; 72392 MW; 118ADD5B53D9D67F CRC64;

Qy	13	MGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDV	72
Db	8	VSIPMSKRNTNGLPGSSSNELKTSAGGAVLSFHDICYRVKVKSGFLFCRKTVEKEILTNI	67
Qy	73	SLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYV	132
Db	68	NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPHG-LSGDVLING-APRPANFKCNSGYV	124
Qy	133	LQSDTLLSSLTVRETLHYTALLAIRRGNPGSF-----QKKVEAVMAELSLSHVADRLIGN	187
Db	125	VQDDVVMGTLTVRENLFQSAALRL---PTTMTNHEKNERINMVIQELGLDKVADSKVGT	180
Qy	188	YSLGGISTGERRRVSIAAQLLDQPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVL	247
Db	181	QFIRGVSGGERKRTSIAMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIF	240
Qy	248	TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCCGYPCPEHSNPFDFYMDLTSV	307
Db	241	SIHQPRYSIFKLFDSLTLASGRLMFHGPAREALGYFASIGYNCEPYNPNADFFLDVING	300
Qy	308	DTQ-----SKEREIETSKRVQMIE---SAYKKSACHKTLKNIE-----RMK	346
Db	301	DSSAVVLSRADRDEGAQEPEEPPEKDTPLIDKLAIFYTNSSFFKDTKVELDQFSGGRKKK	360
Qy	347	HLKTLPMVPFKTKDSPGVFSKLGVLRLRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFV	404
Db	361	KSSVYKEVTTYTT----SFCHQLRWISRRSFKNLLGNPQASVAQIIVTIIILGLVIGAI FYD	416
Qy	405	LRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLPVLRVAVSDQESQDGLYQKWQMM	464
Db	417	LK---NDPSG-IQNRAVLFL-FLTTNQCFSSVSAVELLVVEKKLFIHEYISGYRRVSSYF	471

Qy 465 LAYAL-HVLPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGI 523  
 | :|| :: :||: : |: ||| | | | : : : : | |  
 Db 472 FGKLLSDDLPMRMLPSIIFTCITYFLLGLKPAVGSGFFIMMFTLM--MVAYSASSMALAI 528  
 Qy 524 VQNPNIIVNSVALLSIAGV--LVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFY 581  
 ::|: |::|: | :: || | |:: : : ||: :| | |||  
 Db 529 AAGQSVVSVATLLMTISFVFMFISGLLVNLKTVVPWLSWLQYFSIPRYGFSALQYNEFL 588  
 Qy 582 GLNFTCGSSNVSVTTNPMCAFT--QGIQFIE 610  
 | || | ::||| |:| | :::|  
 Db 589 GQNFCEPG---LNVTTNNTCSFAICTGAEYLE 616

RESULT 10

Q7TMS5

ID Q7TMS5 PRELIMINARY; PRT; 657 AA.  
 AC Q7TMS5;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;  
 RA Strausberg R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC053730; AAH53730.1; -.  
 KW ATP-binding.  
 SQ SEQUENCE 657 AA; 72977 MW; DCD70C5D9FA2BA5F CRC64;

Query Match 19.9%; Score 663; DB 11; Length 657;  
 Best Local Similarity 28.2%; Pred. No. 1.9e-40;  
 Matches 182; Conservative 135; Mismatches 241; Indels 88; Gaps 19;

```

Qy      13 MGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDV 72
      | : | | : | | | | | : | | : : : | : : | | :
Db      12 MSQRNNNGLPRTNSRAVRTLAEGDVLSEFHHITYRV--KVKSGFLV---RKTVEKEILSDI 66

Qy      73 SLYVESGQIMCILGSSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYV 132
      : : : | : | | | : | | : | | | : | | : | | : | |
Db      67 NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPKG-LSGDVLING-APQPAHFKCCSGYV 123

Qy     133 LQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQ-----KKVEAVMAELSLSHVADRLIGN 187
      : | | : : | | | | | : | | : | : : : : | | | | : |
Db     124 VQDDVVMGTLTVRENLFSAALRL----PTTMKNHEKNERINTIIEKLGLEKVADSKVGT 179

Qy     188 YSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGLDCMTANQIVVLLVELARRNRIVVL 247
      : | | | | | : | | : | : | : : | | | | | | : : : | :
Db     180 QFIRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIF 239

Qy     248 TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTSV 307
      : | | | | : | : | | : | : | : | : | : | : | : | :
Db     240 SIHQPRYSIFKLFDSLTLASGKLVFHGPAQKALEYFASAGYHCEPYNNPADFFLDVING 299

Qy     308 DTQS-----KEREIETSKR-----VQMIESAYKKSACHKTLKNIERMKHLKTL 352
      | : : : : | : : | : | : : : | | | : | : : :
Db     300 DSSAVMLNREEQDNEANKTEEPSKGEKPVIENTLSEFYINSAYGETKAELDQL----- 352

Qy     353 MVPFKTKDSPGVFVSKLGV-----LLRRVTRNLVRNKLAVITRLLQNL 394
      | | | | | : | | : | : | | : : | : :
Db     353 -----PGAQEKKGTSAFKEPVYVTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTV 403

Qy     395 IMGLEFL--LFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRAVSDQE 452
      | : | | : : | | : : : : | : | : | : : : | | | : : |
Db     404 ILGLIIGAIYFDLKYDA----AGMQNRAGVLF-FLT TNQCFSSVSAVELFVVEKKLFIHE 458

Qy     453 SQDGLYQKWQMMLAYAL-HVLPFSVVATMIFSSVCYWTGLGHPEVARFGYFSAALLAPHL 511
      | | : : : | | : : | : | | | | | | : : :
Db     459 YISGYRVSSSYFFGKVMSDLPLMRFLPSVIFTCLVLYFMLGLKKTVDFAFFIMFTLI---M 515

Qy     512 IGEFLTLVLLGIVQNPNI VNSVALLSIAGV--LVGSGFLRNIQEMPIPFKIISYFTFQK 569
      : : : | | : : | : | | : : | | | : : : : | : :
Db     516 VAYTASSMALAIATGQSVSVVATLLMTIAFVFMMLFSGLLVNLRTIGPWLSWLQYFSIPR 575

Qy     570 YCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPG 615
      | | | | | | : | | | | : | |
Db     576 YGFTALQYNEFLGQEFCEPG---FNVTDNSTCVNSYAI-----CTG 612

```

RESULT 11

Q8T691

ID Q8T691 PRELIMINARY; PRT; 801 AA.  
 AC Q8T691;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

DE     ABC transporter AbcG1.
GN     ABCG1.
OS     Dictyostelium discoideum (Slime mold).
OC     Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX     NCBI_TaxID=44689;
RN     [1]
RP     SEQUENCE FROM N.A.
RC     STRAIN=Ax4;
RA     Anjard C., Loomis W.F.;
RT     "Evolution of the ABC transporters of Dictyostelium.";
RL     Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC     -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR     EMBL; AF482380; AAL91485.1; -.
DR     GO; GO:0016020; C:membrane; IEA.
DR     GO; GO:0005524; F:ATP binding; IEA.
DR     GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR     GO; GO:0000166; F:nucleotide binding; IEA.
DR     GO; GO:0006810; P:transport; IEA.
DR     InterPro; IPR003593; AAA_ATPase.
DR     InterPro; IPR003439; ABC_transporter.
DR     Pfam; PF00005; ABC_tran; 1.
DR     ProDom; PD000006; ABC_transporter; 1.
DR     SMART; SM00382; AAA; 1.
DR     PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR     PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW     ATP-binding; Transport.
SQ     SEQUENCE 801 AA; 90052 MW; CCC4F0036CB195A3 CRC64;

```

Qy	61	RQQWTRQILKDVSLYVESGQIMCIIIGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRAL	120
		:::     :::        :    :      ::     :   :	
Db	131	KKKISKQILTNINGHIESGTIFAIMGPSGAGKTTLDDILAHLNINGS--GTMYLNGNKS	188
Qy	121	RREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGNPGSFQ-KKVEAVMAELSLSH	179
		:       :       ::     :     : :: : ::  :  :	
Db	189	DFNIFKKLCGYVTQSDSLMPSLTVRETINFYAQLKMPRDVP LKEKLQRVQDIIDEMGLNR	248
Qy	180	VADRLIG--NYSLGISTGERRRVSIAAQLLQDPKVM LFDEPTTGLDCMTANQIVVLLVE	237
		:  : :          : : ::      :     :      : ::   :	
Db	249	CADTLVGTADNKIRGISGGERRRV TISIELLTGPSVILLDEPTSGLDASTSFYVMSALKK	308
Qy	238	LARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFND CGYPCPEHSNP	297
		:   ::       :::    : :     :   :  :         :	
Db	309	LAKSGRTIICTIHQPRSNIYDMFDNLLLLGDGNTIYYGKANKALEYFNANGYHCSEKTNP	368
Qy	298	FDFYMDL--TSVDTS-----	311
		::      : :	
Db	369	ADFFLDLINTQVEDQADSDDDDYNDEEEEIGGGGGSGGGAGGIEDIGISISPTMN GSAV	428
Qy	312	---KER EIE-----TSKR VQMIESAYKKS---AICHKTLKN	341
		:    : : :::     :	
Db	429	DNIKNNELKQQQQQQQQQQST DGRARRRIK KLTKEEMVILKKEYPNSEQGLRVNETLDN	488
Qy	342	I ER-----MKHLKTLPMVPFKTKDSPGVFSKLGVLRRVRTNLVRNKLAVITRLLONLIM	396

```

      | :      | : ||      |      : : || |      | | : :      | : | :
Db      489 ISKENRTDFKYEKT-----RGPNFLTQFSLLLGREVTNAKRHPMAFKVNLIQAIFQ 539

Qy      397 GLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPPVLRRAVSDQESQDG 456
      | || : : : : : || | | : : : : : || : : : : || : : : : |
Db      540 G--LLCGIVYYQLGLGQSSVQSRTGVVAFIIMGVSFPAVMSTIHVFPDVTIFLKDRASG 597

Qy      457 LYQKWQMMLAYALHVLPPFSVVATMIFSSVCYWTLG--LHPEVARFGYFSAALLAPHLIGE 514
      : |      || :      : | : : : ||      : | :      : |      | :
Db      598 VYDTLPFFLAKSFMDACIAVLLPMVTATIVYWMTNQVRDPFYSAAPFFRFVLM---LVLA 654

Qy      515 FLTLVLLGIVQN---PNI-VNSVVALLSIAGVLVSGFLRNIQEMPIPKIISYFTFQKY 570
      | : || : :      || : | : || | :      : || | | : : |      | : | : |
Db      655 SQTCLSLGVLISSSVPNVQVGTAVAPLIVILFFLFSGFFINLNDVPGWLWVFPYISFFRY 714

Qy      571 CSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNFLLILYSF 630
      | | : | | : : || | |      : |      | ||      | | || :
Db      715 MIEAAVINAFKDVHFTCTDSQ---KIGGVCPVQYGNNVIE-NMGYDIDHFWRNWILVLY 770

Qy      631 IPALVILGIVVFKIR 645
      |      : |      : | : :
Db      771 IIGFRVLTFLVLKLK 785

```

RESULT 12

Q9R004

```

ID   Q9R004      PRELIMINARY;      PRT;      657 AA.
AC   Q9R004;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Breast cancer resistance protein 1.
GN   ABCG2 OR BCRP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB; TISSUE=Liver;
RX   MEDLINE=99413474; PubMed=10485464;
RA   Allen J.D., Brinkhuis R.F., Wijnholds J., Schinkel A.H.;
RT   "The mouse Bcrp1/Mxr/Abcp gene: amplification and overexpression in
RT   cell lines selected for resistance to topotecan, mitoxantrone, or
RT   doxorubicin.";
RL   Cancer Res. 59:4237-4241(1999).
DR   EMBL; AF140218; AAD54216.1; -.
DR   MGD; MGI:1347061; Abcg2.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0005524; F:ATP binding; IEA.
DR   GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR   GO; GO:0000166; F:nucleotide binding; IEA.
DR   GO; GO:0006810; P:transport; IEA.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transporter.
DR   InterPro; IPR006162; Ppantne_S.
DR   Pfam; PF00005; ABC_tran; 1.

```



DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 1.  
KW ATP-binding.  
SQ SEQUENCE 657 AA; 73021 MW; 207B70BC272CC0D5 CRC64;

Query Match 19.8%; Score 660; DB 11; Length 657;  
Best Local Similarity 28.0%; Pred. No. 3.1e-40;  
Matches 181; Conservative 135; Mismatches 242; Indels 88; Gaps 19;

Qy 13 MGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDV 72  
| : | | | | | | | | : | : : : | : : | | :  
Db 12 MSQRNNNGLPRMNSRAVRTLAEGDVLSFHHITYRV--KVKSGFLV---RKTVEKEILSDI 66

Qy 73 SLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYV 132  
: : : | : | | : | | : | | : | | : | : | | :  
Db 67 NGIMKPG-LNAILGPTGGGKSSLLDVLAARKDPKG-LSGDVLING-APQPAHFKCCSGYV 123

Qy 133 LQSDTLLSSLTVRETLHYTALLAIRRGNGPSFQ-----KKVEAVMAELSLSHVADRLIGN 187  
: | | : : | | | | | : | : | : : : : | | | : |  
Db 124 VQDDVVMGTLTVRENLFQFSAALRL----PTTMKNHEKNERINTIIEKLGLEKVADSKVGT 179

Qy 188 YSLGGISTGERRRVSIQAQLLQDPKVMFLFDEPTTGLDCMTANQIVVLLVELARRNRIVVL 247  
: | | | | : | | : | : | | : | | | | | : : : | : :  
Db 180 QFIRGISGGERKRTSIGMELITDPSILFLDEPTTGLDSSSTANAVLLLLKRMSKQGRTIIF 239

Qy 248 TIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPFDFYMDLTSV 307  
: | | | | : | : | | : : : | : | | : | : | : : :  
Db 240 SIHQPRYSIFKLFDLSLTLLASGKLVFHGPAQKALEYFASAGYHCEPYNPADFFLDVING 299

Qy 308 DTQS-----KEREIETSKR-----VQMIESAYKKSACHKTLKNIERMKHLKTL 352  
| : : : | : : | : : : : : | : | : : : : :  
Db 300 DSSAVMLNREEQDNEANKTEEPSKGEKPVIENTLSEFYINSAIYGETKAEILDQL----- 352

Qy 353 MVPFKTKDSPGVFSKLGV-----LLRRVTRNLVRNKLAVITRLLQNL 394  
| | | | : | | : | : | | : : : : :  
Db 353 -----PGAQEKKGTSAFKEPVYVTSFCHQLRWIARRSFKNLLGNPQASVAQLIVTV 403

Qy 395 IMGLFL--LFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQE 452  
| : | | : : | | : : : : | : | : : : : : | : : |  
Db 404 ILGLIIGAIYFDLKYDA---AGMQNRAGVLF-FLT TNQCFSSVSAVELEFVVEKKLFIHE 458

Qy 453 SQDGLYQKWQMMLAYAL-HVLPFSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHL 511  
| | : : : | : : : : | : | | | | : : :  
Db 459 YISGYRVSSYFFGKVMSDLPLMRFLPSVIFTCILYFMLGLKKTVD AFFIMFTLI---M 515

Qy 512 IGEFLTLVLLGIVQNPNI VNSVVALLSIAGV--LVGSGFLRN IQEMPIPKIISYFTFQK 569  
: : : | | : : : | : | | : : | | : : : : : :  
Db 516 VAYTASSMALAIATGQSVSVATLLMTIAFVFMMLFSGLLVNLRITIGPWL SWLQYFSIPR 575

Qy 570 YCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPG 615  
| | | | | : | | | : | | :  
Db 576 YGFTALQYNEFLGQEFCEPG---FNVTDNSTCVNSYAI-----CTG 612

RESULT 13

Q9M3D6

ID Q9M3D6 PRELIMINARY; PRT; 725 AA.  
AC Q9M3D6;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE ABC transporter-like protein.  
GN T26I12.10 OR AT3G55130.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Monfort A., Casacuberta E., Puigdomenech P., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X., Quetier F., Salanoubat M.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,  
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,  
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,  
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,  
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,  
RA Ecker J.R., Theologis A.;  
RT "Full Length cDNA of gene T26I12.10/AT3g55130 (GI:7019646).";  
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,  
RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,  
RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,  
RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;  
RT "Arabidopsis Open Reading Frame (ORF) Clones.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AL132954; CAB75747.1; -.  
DR EMBL; AY045932; AAK76606.1; -.  
DR EMBL; AY079387; AAL85118.1; -.  
DR PIR; T47652; T47652.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.  
DR GO; GO:0000166; F:nucleotide binding; IEA.  
DR GO; GO:0006810; P:transport; IEA.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.

DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW ATP-binding.  
SQ SEQUENCE 725 AA; 80656 MW; 790C535A7929CC16 CRC64;

Query Match 19.5%; Score 649.5; DB 10; Length 725;  
Best Local Similarity 29.4%; Pred. No. 2.1e-39;  
Matches 182; Conservative 124; Mismatches 246; Indels 67; Gaps 15;

```
Qy      33 PEPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVS LYVESGQIMCILGSSGSGK 92
      | | : | : | | : | | | | : : | | | | | : : | | : | | : | |
Db      68 PVPYVLNFNNLQYDVTLR--FGFSRQNGVKTLDDVSGEASGDILAVLGASGAGK 123

Qy      93 TTLDDAMSGRLGRAGTFLGEVYVNG-RALRREQFQDCFSYVLQSDTLLSSLTVRETLYHT 151
      : | | : | | : | : | : | : | : | : | : | : | : | : | : | : |
Db     124 STLIDALAGRAE-GSLRGSVTLNGEKLQSRLLKVISAYVMQDDLLFPMLTVKETLMFA 182

Qy     152 ALLAIRRG-NPGSFQKKVEAVMAELSLSHVADRLIGNYSLGISTGERRRVSIAAQLIQD 210
      : : | : : | | : | : | : | : | : | : | : | : | : | : | : |
Db     183 SEFRLPRSLSKSKKMERVEALIDQLGLRNAANTVIGDEGHRGVSGGERRRVSIGIDIHD 242

Qy     211 PKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFE 270
      | | : | | : | | | : | : | : | : | : | : | : | : | : | : |
Db     243 PIVLFLDEPTSGLDSTNAFMVVQVLKRIAQSGSIVIMSIHQPSARIVELLDRLIILSRGK 302

Qy     271 LIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETS----- 319
      : | | : | : | | | | | : | : | | | : | : | | | : | : |
Db     303 SVENGSPASLPGFFSDFGRPIPEKENISEFALDLV-----RELEGSNEGTKALVDFN 354

Qy     320 -----KRVQMIESAYK-----KSAICHKTL--KNIERMKHLKTLPMVPFKTKD 360
      : : : | : | : | | | | | : | : | | | : | : | | | : | : |
Db     355 EKWQONKISLIQSAPQTNKLDQDRSLSLKEAINASVSRGKLVSGSSRSNPTSMETVSSYA 414

Qy     361 SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV 420
      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     415 NPSLFETF-ILAKRYMKNWIRMPVLGTRIATVMVTGC-LLATVYWKLDHTPRGA-QERL 471

Qy     421 GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATM 480
      | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db     472 -TLFAFVVPTMFYCCLDNVPVFIQERYIFLRETHNAYRTSSYVISHSLVSLPQLLAPSL 530

Qy     481 IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIVNS-VVALLSI 539
      : | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     531 VFSAITFWTVGLSGGLEGFVYCLLIYASFWSGSSSVVTFISGVV--PNIMLCYMSITYL 588

Qy     540 AGVLVGSGFLRNIQEMPIPFKII SYFTFQKYCSEILVNEFYGLNFTCGSSNVSVTTNPM 599
      | | : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db     589 AYCLLLSGFYVNRDRI PFYWTWFHYISILKYPYEAVLINEF-----DDPS 633

Qy     600 CAFTQGIQFIEKTCPGATS 618
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     634 RCFVRGVQVFDSTLLGGVS 652
```

RESULT 14

Q7YYX5

ID Q7YYX5 PRELIMINARY; PRT; 643 AA.

AC Q7YYX5;



```

      | :|| : | :      :      : : : : |      | : || : : :
Db      477 VPILVFESHIFYFMSNTNSVSPGWNTLTQYLCYQLTILLTSWASYGLVYFICGITKSLEL 536

Qy      530 VNSVVALLSIAGVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGS 589
      : | : | | :| ||| : : :| : | : || :| |||| | | |
Db      537 AYGIAPLIIFVIV-SGFYVTVNKLPLWVSWIKYISFQRYSSALVVNTF-PPNQNWGP 594

Qy      590 SNVSVTTNPMCAFTQGIQF-IEKTCPGATSRFTMNFILILYSFIPALVILGIV 640
      :      || | :| | :| : :      || : || :
Db      595 IQTDILLK-----QFSIDQT-----SFLNNAV-----LVVLGIL 624

```

RESULT 15

Q8RWI9

```

ID      Q8RWI9      PRELIMINARY;      PRT;      691 AA.
AC      Q8RWI9;
DT      01-JUN-2002 (TrEMBLrel. 21, Created)
DT      01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Hypothetical protein.
GN      AT3G21090.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA      Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA      Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA      Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA      Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA      Ecker J., Theologis A., Davis R.W.;
RL      Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RA      Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,
RA      Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA      Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA      Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA      Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA      Ecker J., Theologis A., Davis R.W.;
RL      Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
CC      -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR      EMBL; AY093054; AAM13053.1; -.
DR      EMBL; BT000405; AAN15724.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0005524; F:ATP binding; IEA.
DR      GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; IEA.
DR      GO; GO:0000166; F:nucleotide binding; IEA.
DR      GO; GO:0006810; P:transport; IEA.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.

```

DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Hypothetical protein; ATP-binding; Transport.  
 SQ SEQUENCE 691 AA; 77219 MW; CE473CC0B440D7E9 CRC64;

Query Match 18.9%; Score 628; DB 10; Length 691;  
 Best Local Similarity 28.1%; Pred. No. 7.6e-38;  
 Matches 173; Conservative 123; Mismatches 225; Indels 94; Gaps 17;

```

Qy      25 LEGAPATAPE-PHSLGILHASYSVSHRVRPWWDDITSCRQQW----TRQILKDVSLYVESG 79
      |||: : : | | : :| | | : | | : :| | : :| |
Db      3 LEGSSSGRRQLPSKLEMSRGAYLA-----WEDLTVVIPNFSDGPTRRLQLRLNGYAEFG 56

Qy      80 QIMCILGSSSGSKTTLDDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLL 139
      :|| |:| ||||:||||:|||| | | : :||: | : :|| | | |
Db      57 RIMAIMGPSGSGKSTLLDSLARGRLARNVVMGTGNLLNGKKARLD--YGLVAYVTQEDVLL 114

Qy     140 SSLTVRETLYHTALLAIRGNPGSFQKK-----VEAVMAELSLSHVADRIGNYSLGGIS 194
      :|||||: |:| | : | | : | | : || | : ||:||||: |:|
Db     115 GTLTVRETITYSAHLRL----PSDMSKEEVSDIVEGTIIEGLQDCSDRVIGNWHARGVS 170

Qy     195 TGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRS 254
      |||:||||| :| |:: | ||:| | :| : : | :|| | | : :|| |
Db     171 GGERKRVSIALEILTRPQILFLDEPTSGLDASAFFVIQALRNRIARDGRTVISSVHQPSS 230

Qy     255 ELFQLEDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDFYMDLTSVDTQSKER 314
      |:| ||| : :|| || : : | :|| : |:| | : || : | : :
Db     231 EVFALFDDLFLSSGESVYFGEAKSAVEFFAESGFPCPKRNPDSHFLRCINSDFDFTVTA 290

Qy     315 EIETSKRVQ-----MIESAYKKSACHKTLKNIERMKHLKTLPMV 354
      : : |:| | : :||: ||: | | : : : | |
Db     291 TLKGSQRIQETPATSDPLMNLATSVIKARLVEN-YKRSKYAKSAKSRIRELSNIEGLEME 349

Qy     355 PFKTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLMGLFLLFFVLRVRSNVLKG 414
      | : : : :| | | | : | : | | : : : :
Db     350 IRKGSEATW-WKQLRSLTARSFINMCRDVGYYWTRIISYIVVSI----- 392

Qy     415 AIQDRVGLLYQFVGATPYTGMLNAVNL-----FPVL---RAVSDQESQDG 456
      || : : || : || : | | : | | | : | |
Db     393 ----SVGTIFYDVGY-SYTSILARVSCGGFITGFMTFMSIGGFPSFLEEMKVYKERLSG 447

Qy     457 LYQKWQMMLAYALHVLFPFSVATMIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEF 516
      | : : : || | :| : : | : | : : :| : : :| |
Db     448 YYGVSUYILSNYISSFPFLVAISVITGTITYNLVKFRPGFSHYAFFCLNIFFSVSVIESL 507

Qy     517 TLVLLGIVQNPNIIVNSVALLSIAG-VLVGSGFLRNIEQEMPIPKI-----ISYFTFQKY 570
      :|: :| || : : : | : : || | : :| || :|| : : :
Db     508 MMVVASVV--PNFLMGLITAGLIGIIMMTSGFFRLLPDLP---KIFWRYPVSYISYGSW 562

Qy     571 CSEILVVNEFYGLNF 585
      : | : | | |
Db     563 AIQGGYKNDFLGLF 577
  
```

Search completed: February 27, 2004, 07:15:28  
 Job time : 37.1394 secs

OM protein - protein search, using sw model

Run on: February 27, 2004, 06:40:43 ; Search time 10.0797 Seconds  
(without alignments)  
3362.970 Million cell updates/sec

Title: US-09-989-981A-6  
Perfect score: 3326  
Sequence: 1 MGDLSLTPGGSMGLQVNRG.....PALVILGIVVFKIRDHLISR 651

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	3326	100.0	651	1	ABG5_HUMAN	Q9h222 homo sapien
2	2738.5	82.3	652	1	ABG5_MOUSE	Q99pe8 mus musculu
3	2721.5	81.8	652	1	ABG5_RAT	Q99pe7 rattus norv
4	698	21.0	673	1	ABG8_MOUSE	Q9dbm0 mus musculu
5	697	21.0	673	1	ABG8_HUMAN	Q9h221 homo sapien
6	692.5	20.8	694	1	ABG8_RAT	P58428 rattus norv
7	676.5	20.3	655	1	ABG2_HUMAN	Q9unq0 homo sapien
8	627	18.9	1294	1	YOH5_YEAST	Q08234 saccharomyc
9	623	18.7	677	1	WHIT_LUCCU	Q05360 lucilia cup
10	621	18.7	1049	1	ADP1_YEAST	P25371 saccharomyc
11	607.5	18.3	695	1	WHIT_ANOGA	Q27256 anopheles g
12	602.5	18.1	687	1	WHIT_DROME	P10090 drosophila
13	596.5	17.9	678	1	ABG1_HUMAN	P45844 homo sapien
14	591	17.8	679	1	WHIT_CERCA	Q17320 ceratitidis c
15	578.5	17.4	646	1	ABG4_HUMAN	Q9h172 homo sapien
16	573	17.2	709	1	WHIT_ANOAL	Q16928 anopheles a
17	569.5	17.1	666	1	ABG1_MOUSE	Q64343 mus musculu

18	561.5	16.9	598	1	YPC3_CAEEL	Q11180	caenorhabdi
19	547	16.4	666	1	SCRT_DROME	P45843	drosophila
20	517.5	15.6	610	1	YQ5C_CAEEL	Q09466	caenorhabdi
21	454.5	13.7	675	1	BROW_DROME	P12428	drosophila
22	435	13.1	668	1	BROW_DROVI	Q24739	drosophila
23	427	12.8	1499	1	CDR2_CANAL	P78595	candida alb
24	406.5	12.2	1333	1	YN99_YEAST	P53756	saccharomyc
25	403.5	12.1	1564	1	PDRA_YEAST	P51533	saccharomyc
26	399.5	12.0	1530	1	BFR1_SCHPO	P41820	schizosacch
27	398	12.0	1501	1	SNQ2_YEAST	P32568	saccharomyc
28	392.5	11.8	1501	1	CDR3_CANAL	O42690	candida alb
29	391	11.8	1529	1	PDRF_YEAST	Q04182	saccharomyc
30	383.5	11.5	650	1	ABG3_MOUSE	Q99p81	mus musculu
31	382	11.5	1501	1	CDR1_CANAL	P43071	candida alb
32	382	11.5	1511	1	PDRC_YEAST	Q02785	saccharomyc
33	374	11.2	1511	1	PDR5_YEAST	P33302	saccharomyc
34	367	11.0	1410	1	PDRB_YEAST	P40550	saccharomyc
35	365.5	11.0	1490	1	CDR4_CANAL	O74676	candida alb
36	263.5	7.9	345	1	METN_HAEIN	P44785	haemophilus
37	263	7.9	241	1	YHBG_HAEIN	P45073	haemophilus
38	257	7.7	365	1	CYSA_ECO57	Q8xbj8	escherichia
39	257	7.7	365	1	CYSA_ECOL6	Q8ffb3	escherichia
40	257	7.7	365	1	CYSA_ECOLI	P16676	escherichia
41	254	7.6	364	1	CYSA_SALTI	Q8z4v6	salmonella
42	254	7.6	365	1	CYSA_SALTY	P40860	salmonella
43	251.5	7.6	363	1	CYSA_YERPE	Q8d0w8	yersinia pe
44	251	7.5	573	1	CYDC_ECOLI	P23886	escherichia
45	250.5	7.5	231	1	YTRE_BACSU	O34392	bacillus su

# ALIGNMENTS

## RESULT 1

### ABG5\_HUMAN

ID ABG5\_HUMAN STANDARD; PRT; 651 AA.  
AC Q9H222;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).  
GN ABCG5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND VARIANT GLU-604.  
RC TISSUE=Liver;  
RX MEDLINE=20553648; PubMed=11099417;  
RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,  
RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;  
RT "Accumulation of dietary cholesterol in sitosterolemia caused by  
RT mutations in adjacent ABC transporters."  
RL Science 290:1771-1775(2000).  
RN [2]  
RP SEQUENCE FROM N.A., VARIANTS SITOSTEROLEMIA HIS-389; HIS-419 AND



RP PRO-419, AND VARIANT GLU-604.  
 RC TISSUE=Liver;  
 RX MEDLINE=20578753; PubMed=11138003;  
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,  
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,  
 RA Dean M., Patel S.B.;  
 RT "Identification of a gene, ABCG5, important in the regulation of  
 RT dietary cholesterol absorption."  
 RL Nat. Genet. 27:79-83(2001).  
 RN [3]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism."  
 RL J. Lipid Res. 42:1513-1520(2001).  
 RN [4]  
 RP VARIANTS SITOSTEROLEMIA GLN-146; HIS-389; PRO-419; HIS-419 AND  
 RP SER-550, AND VARIANT GLU-604.  
 RX MEDLINE=21344600; PubMed=11452359;  
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,  
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,  
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,  
 RA Patel S.B.;  
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic  
 RT structure and spectrum of mutations involving sterolin-1 and  
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively."  
 RL Am. J. Hum. Genet. 69:278-290(2001).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
 CC ABCG8 along a pathway regulating dietary-sterol absorption and  
 CC excretion.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels  
 CC in the small intestine and colon.  
 CC -!- DISEASE: Defects in ABCG5 are a cause of sitosterolemia  
 CC [MIM:210250]; also known as phytosterolemia or shellfish  
 CC sterolemia. It is a rare autosomal recessive disorder  
 CC characterized by increased intestinal absorption of all sterols  
 CC including cholesterol, plant and shellfish sterols, and decreased  
 CC biliary excretion of dietary sterols into bile. Sitosterolemia  
 CC patients have hypercholesterolemia, very high levels of plant  
 CC sterols in the plasma, and frequently develop tendon and tuberous  
 CC xanthomas, accelerated atherosclerosis and premature coronary  
 CC artery disease.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).



Qy	121	RREQFQDCFSYVLQSDTLSSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Db	121	RREQFQDCFSYVLQSDTLSSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSHV	180
Qy	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Db	181	ADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR	240
Qy	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Db	241	RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF	300
Qy	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Db	301	YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD	360
Qy	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Db	361	SPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRV	420
Qy	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Db	421	GLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATM	480
Qy	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Db	481	IFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSIA	540
Qy	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Db	541	GVLVGSGFLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMC	600
Qy	601	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	651
Db	601	AFTQGIQFIEKTCPGATSRFTMNFILILYSFIPALVILGIVVFKIRDHLISR	651

# RESULT 2

## ABG5\_MOUSE

ID ABG5\_MOUSE STANDARD; PRT; 652 AA.

AC Q99PE8;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).

GN ABCG5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Liver;

RX MEDLINE=20578753; PubMed=11138003;

RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,

RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,

RA Dean M., Patel S.B.;  
 RT "Identification of a gene, ABCG5, important in the regulation of  
 RT dietary cholesterol absorption.";  
 RL Nat. Genet. 27:79-83(2001).  
 RN [2]  
 RP TISSUE SPECIFICITY, AND INDUCTION.  
 RX MEDLINE=20553648; PubMed=11099417;  
 RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,  
 RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;  
 RT "Accumulation of dietary cholesterol in sitosterolemia caused by  
 RT mutations in adjacent ABC transporters.";  
 RL Science 290:1771-1775(2000).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
 CC ABCG8 along a pathway regulating dietary-sterol absorption and  
 CC excretion (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower  
 CC level, in the liver.  
 CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated  
 CC by the liver X receptor/retinoic X receptor (LXR/RXR) pathway.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF312713; AAG53097.1; -.  
 DR MGD; MGI:1351659; Abcg5.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Glycoprotein; Transmembrane; Transport.  
 FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 386 406 1 (POTENTIAL).  
 FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 423 443 2 (POTENTIAL).  
 FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 464 484 3 (POTENTIAL).  
 FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 505 525 4 (POTENTIAL).  
 FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 530 550 5 (POTENTIAL).  
 FT DOMAIN 551 622 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 623 643 6 (POTENTIAL).

FT DOMAIN 644 652 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 87 94 ATP (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 652 AA; 73244 MW; 80CE37ADCC19771E CRC64;

Query Match 82.3%; Score 2738.5; DB 1; Length 652;  
 Best Local Similarity 80.1%; Pred. No. 2.1e-188;  
 Matches 522; Conservative 64; Mismatches 65; Indels 1; Gaps 1;

Qy 1 MGDLSSTLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59  
 ||:| |:| |:| :||| ||| | | |||:| | |||:| | ||:| |  
 Db 1 MGELPFLSPEGARGPHINRGSLSSLEQGSVTGTEARHSLGVLHVSYSVSNRVGPWWNIKS 60  
  
 Qy 60 CRQQWTRQILKDVSLYVESGQIMCILGSSSGSKTTLDDAMSGRLGRAGTFLGEVYVNGRA 119  
 |:|:| ||||| |:|:| ||||| |:|:| ||||| |:|:| ||||| |:|:| |||||  
 Db 61 CQQKWDRQILKDVSLYIESGQIMCILGSSSGSKTTLDDAISGRLRRTGTLEGEVVFVNGCE 120  
  
 Qy 120 LRREQFQDCFSYVLQSDTLSSSLTVRETLYHTALLAIRRGNPGSFQKKVEAVMAELSLSH 179  
 |||:||||| ||||| ||||| |||:|:| : : ||||| |||||  
 Db 121 LRRDQFQDCFSYVLQSDVFLSSSLTVRETLYRTAMLALCRSSADFYNNKVEAVMTLSLSH 180  
  
 Qy 180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMLEFDEPTTGLDCMTANQIVVLLVELA 239  
 |||:|:|:|:| ||||:||||| |||||:| |||||:||||| |||||:| |||  
 Db 181 VADQMIGSYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANQIVLLLAELA 240  
  
 Qy 240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF 299  
 ||:|||:|:||||| |||||:|:|:|:| ||| |||:||||| |||||  
 Db 241 RRDRIVIVTIHQPRSELFQHFQDKIAILTYGELVFCGTPEEMLGFFNCCGYPCPEHSNPF 300  
  
 Qy 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPF 359  
 ||||| |||||:||||| |||||:| |:|:| | | |:||| :|:|||| | |||||  
 Db 301 FYMDLTSVDTQSREREIETYKRVQMLECAFKESDIYHKILENIERARYLKTLPVPF 360  
  
 Qy 360 DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419  
 | ||:| ||||| |||||:||| ||| ||:||||| |:|:|:|:| ||||:| |||  
 Db 361 DPPGMFGKLGVLRRVTRNLVRNKLAVIMRLVQNLIMGLFLIFYLLRVQNTLKGAVQDR 420  
  
 Qy 420 VGLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQDGLYQKWQMLLAYALHVLFPFSV 479  
 ||||| ||||| |||||:||||| ||||| ||||:| | |||||:| |  
 Db 421 VGLLYQLVGATPYTGMLNAVNLFPMLRAVSDQESQDGLYHKWQMLLAYVLHVLFPFSV 480  
  
 Qy 480 MIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNI VNSV 539  
 :||||| |||||:||||| ||||| ||||| ||||| |||||:|||||  
 Db 481 VIFSSVCYWTGLGLYPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNI VNSI 540  
  
 Qy 540 AGVLVSGSGLRNIQEMPIPFKIIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSV 599  
 |:|:|:|:|:| ||||| ||:| ||||| ||||| ||||| ||| |:|:|  
 Db 541 SGLLIGSGFIRNIQEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGGSNTSML 600  
  
 Qy 600 CAFTQGIQFIEKTCPGATSRTMNFILILYSFIPALVILGIVVFKIRDHLISR 651  
 || |||:||||| ||||| ||||| |||||:|:|:|:| |||  
 Db 601 CAITQGVQFIEKTCPGATSRTANFLILYGFIPALVILGIVIFKVRDYILISR 652

RESULT 3

ABG5\_RAT  
 ID ABG5\_RAT STANDARD; PRT; 652 AA.  
 AC Q99PE7; Q8CIQ4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 5 (Sterolin-1).  
 GN ABCG5.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Small intestine;  
 RX MEDLINE=20578753; PubMed=11138003;  
 RA Lee M.-H., Lu K., Hazard S., Yu H., Shulenin S., Hidaka H., Kojima H.,  
 RA Allikmets R., Sakuma N., Pegoraro R., Srivastava A.K., Salen G.,  
 RA Dean M., Patel S.B.;  
 RT "Identification of a gene, ABCG5, important in the regulation of  
 RT dietary cholesterol absorption."  
 RL Nat. Genet. 27:79-83(2001).  
 RN [2]  
 RP REVISION TO 2.  
 RA Lu K., Lee M.-H., Patel S.B.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND VARIANT CYS-583.  
 RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;  
 RX PubMed=12783625;  
 RA Yu H., Pandit B., Klett E., Lee M.H., Lu K., Helou K., Ikeda I.,  
 RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;  
 RT "The rat STSL locus: characterization, chromosomal assignment, and  
 RT genetic variations in sitosterolemic hypertensive rats."  
 RL BMC Cardiovasc. Disord. 3:4-4(2003).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG8 or be tightly coupled to  
 CC ABCG8 along a pathway regulating dietary-sterol absorption and  
 CC excretion (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- TISSUE SPECIFICITY: Expressed only in liver and intestine.  
 CC -!- POLYMORPHISM: The polymorphism at position 583 is found in strains  
 CC SHR, SHRSP and Wistar Kyoto which are both hypertensive and  
 CC sitosterolemic. Strains which are hypertensive but not  
 CC sitosterolemic do not contain a polymorphism at this position.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF312714; AAG53098.3; -.
DR EMBL; AY145899; AAN64275.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Polymorphism.
FT DOMAIN 1 385 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 386 406 1 (POTENTIAL).
FT DOMAIN 407 422 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 423 443 2 (POTENTIAL).
FT DOMAIN 444 463 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 464 484 3 (POTENTIAL).
FT DOMAIN 485 504 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 505 525 4 (POTENTIAL).
FT DOMAIN 526 529 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 530 550 5 (POTENTIAL).
FT DOMAIN 551 624 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 625 645 6 (POTENTIAL).
FT DOMAIN 646 652 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 87 94 ATP (POTENTIAL).
FT CARBOHYD 585 585 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 583 583 G -> C (in strains SHR, SHRSP and Wistar
FT Kyoto).
SQ SEQUENCE 652 AA; 73372 MW; 49FEF7372269299D CRC64;

```

```

Query Match 81.8%; Score 2721.5; DB 1; Length 652;
Best Local Similarity 79.3%; Pred. No. 3.4e-187;
Matches 517; Conservative 68; Mismatches 66; Indels 1; Gaps 1;

```

```

Qy 1 MGDLSLTPGGSMGLQVNRGSQSSLEGAPATAPEP-HSLGILHASYSVSHRVRPWWDITS 59
| :| |:| |:| | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MSELPLFLSPGARGPHNNRGSQSSLEEGSVTGSEARHSLGVLNVSFVSNRVGPWWNIKS 60

Qy 60 CRQQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRA 119
|:|:| |:| | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 CQQKWDRKILKDVSLYIESGQTMCILGSSGSGKTTLLDAISGRLRRTGTLEGEVVFVNGCE 120

Qy 120 LRREQFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLSH 179
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 LRRDQFQDCVSYLLQSDVFLSSLTVRETLRYTAMLALRSSADFYDKKVEAVLTELSLSH 180

Qy 180 VADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLDEPTTGLDCMTANQIVVLLVELA 239
| | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 VADQMIGNYNFGGISSGERRRVSIAAQLLQDPKVMMLDEPTTGLDCMTANHIVLLLVELA 240

Qy 240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPF 299
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 RRNRIVIVTIHQPRSELFHHFDKIAILTYGELVFCGTPPEMLGFFNDCGYPCPEHSNPF 300

Qy 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK 359

```

```

Db      301 FYMDLTSVDTQSREREIETYKRVQMLESAFRQSDICHKILENIERTRHLKTLPMVPFKTK 360
Qy      360 DSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 419
      : ||:| |||||:|||| || ||:|||||:|:|:|:|:|:|:|:|
Db      361 NPPGMFCKLGVLLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDR 420
Qy      420 VGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSVVAT 479
      ||||| |||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      421 VGLLYQLVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMLLAYALHVLPPFSIVAT 480
Qy      480 MIFSSVCYWTLGLHPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSI 539
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db      481 VIFSSVCYWTLGLYPEVARFGYFSAALLAPHLIGEFLLVLLGIVQNPNIIVNSVVALLSI 540
Qy      540 AGVLVSGGFLRNIQEMPIPFKIIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPM 599
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      541 SGLLIGSGFIRNIEEMPIPLKILGYFTFQKYCCEILVVNEFYGLNFTCGSSNTSVPNNPM 600
Qy      600 CAFTQGIQFIEKTCPGATSRTMNFILILYSFIPALVILGIVVFKIRDHLISR 651
      |: |||||:||||| |||||:|||||:|:|:|:|:|
Db      601 CSMTQGIQFIEKTCPGATSRTTNFLILYSFIPTLVILGMVVFVKVRDYILISR 652

```

#### RESULT 4

##### ABG8\_MOUSE

```

ID  ABG8_MOUSE      STANDARD;      PRT;      673 AA.
AC  Q9DBM0;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  ATP-binding cassette, sub-family G, member 8 (Sterolin-2).
GN  ABCG8.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC  STRAIN=C57BL/6; TISSUE=Liver;
RX  MEDLINE=21344600; PubMed=11452359;
RA  Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,
RA  Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,
RA  Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,
RA  Patel S.B.;
RT  "Two genes that map to the STSL locus cause sitosterolemia: genomic
RT  structure and spectrum of mutations involving sterolin-1 and
RT  sterolin-2, encoded by ABCG5 and ABCG8, respectively.";
RL  Am. J. Hum. Genet. 69:278-290(2001).
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORM 1).
RC  STRAIN=C57BL/6J; TISSUE=Liver;
RX  MEDLINE=21085660; PubMed=11217851;
RA  Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA  Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA  Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA  Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

```



RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzairelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).

RN [3]

RP TISSUE SPECIFICITY, AND INDUCTION.

RX MEDLINE=20553648; PubMed=11099417;

RA Berge K.E., Tian H., Graf G.A., Yu L., Grishin N.V., Schultz J.,  
RA Kwiterovich P., Shan B., Barnes R., Hobbs H.H.;

RT "Accumulation of dietary cholesterol in sitosterolemia caused by  
RT mutations in adjacent ABC transporters.";

RL Science 290:1771-1775(2000).

CC -!- FUNCTION: Transporter that appears to play an indispensable role  
CC in the selective transport of the dietary cholesterol in and out  
CC of the enterocytes and in the selective sterol excretion by the  
CC liver into bile.

CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to  
CC ABCG5 along a pathway regulating dietary-sterol absorption and  
CC excretion (By similarity).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=1;

CC IsoId=Q9DBM0-1; Sequence=Displayed;

CC Name=2;

CC IsoId=Q9DBM0-2; Sequence=VSP\_000053;

CC Note=No experimental confirmation available;

CC -!- TISSUE SPECIFICITY: Expressed in the intestine and, at lower  
CC level, in the liver.

CC -!- INDUCTION: Upregulated by cholesterol feeding. Possibly mediated  
CC by the liver X receptor/retinoic acid X receptor (LXR/RXR) pathway.

CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
CC subfamily.

CC -!- CAUTION: Seems to have a defective ATP-binding region.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AF324495; AAK84079.1; -.

DR EMBL; AK004871; BAB23630.1; -.

```

DR      MGD; MGI:1914720; Abcg8.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      Glycoprotein; Transmembrane; Transport; Alternative splicing.
FT      DOMAIN          1      413      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM        414      434      1 (POTENTIAL).
FT      DOMAIN          435      447      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM        448      468      2 (POTENTIAL).
FT      DOMAIN          469      496      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM        497      517      3 (POTENTIAL).
FT      DOMAIN          518      526      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM        527      547      4 (POTENTIAL).
FT      DOMAIN          548      569      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM        570      590      5 (POTENTIAL).
FT      DOMAIN          591      639      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM        640      660      6 (POTENTIAL).
FT      DOMAIN          661      673      CYTOPLASMIC (POTENTIAL).
FT      CARBOHYD        619      619      N-LINKED (GLCNAC. . .) (POTENTIAL)
FT      VARSPLIC        377      377      Missing (in isoform 2).
FT                                          /FTid=VSP_000053.
SQ      SEQUENCE        673 AA;  75995 MW;  78012611A5DF2589 CRC64;

```

Qy	394	LIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPPYTGMLNAVNLFPVLRVSDQES	453
		:    :   :     :   :   :   :	
Db	425	CLMSLIIGFLYYGHGAKQL--SFMDTAALLFMIGALIPFNVILDVVSCHSERSMLYYEL	482
Qy	454	QDGLYQKWQMMMLAYALHVLPPFSVVATMIFSSVCYWTGLGLHPEVAREFGYFSAALLAPHLIG	513
		:              :  :              :	
Db	483	EDGLYTAGPYFFAKILGELPEHCAYVIIYAMPIYWLTLNLRPVEPELF-----LL--HFL	534
Qy	514	EFL-----TLVLLGIVQNPNI-VNSVVALLSIAGVLVSGFRLNIQEMPIPKIISYFT	566
		:    :     :  :  :  :  :    :	
Db	535	VWLTVFCCRTMALAASAMLPFTFHMSFFCNALYNSFYLTAGFMINLDNLWIVPAWISKLS	594
Qy	567	FQKYCSEILVNEFYGLNFT--CGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF	624
		:    :     :    :   :     :	
Db	595	FLRWCFSGLMQIQFNHLYTTQIGNFTFSILGDTM-----ISAMDLNS	637
Qy	625	LILYSFIPALVILGI	639
		:   :  :	
Db	638	HPLYAIY--LIVIGI	650

RL Am. J. Hum. Genet. 69:278-290(2001).  
 RN [3]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to  
 CC ABCG5 along a pathway regulating dietary-sterol absorption and  
 CC excretion.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9H221-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9H221-2; Sequence=VSP\_000052;  
 CC Note=Minor form detected in approximately 10% of the cDNA  
 CC clones;  
 CC -!- TISSUE SPECIFICITY: Strongly expressed in the liver, lower levels  
 CC in the small intestine and colon. Detectable in a wide variety of  
 CC human tissues.  
 CC -!- DISEASE: Defects in ABCG8 are a cause of sitosterolemia  
 CC [MIM:210250]; also known as phytosterolemia or shellfish  
 CC sterolemia. It is a rare autosomal recessive disorder  
 CC characterized by increased intestinal absorption of all sterols  
 CC including cholesterol, plant and shellfish sterols, and decreased  
 CC biliary excretion of dietary sterols into bile. Sitosterolemia  
 CC patients have hypercholesterolemia, very high levels of plant  
 CC sterols in the plasma, and frequently develop tendon and tuberous  
 CC xanthomas, accelerated atherosclerosis and premature coronary  
 CC artery disease.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -!- CAUTION: Seems to have a defective ATP-binding region.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF320294; AAG40004.1; -.  
 DR EMBL; AF324494; AAK84078.1; -.  
 DR EMBL; AF351824; AAK84663.1; -.  
 DR EMBL; AF351812; AAK84663.1; JOINED.  
 DR EMBL; AF351813; AAK84663.1; JOINED.  
 DR EMBL; AF351814; AAK84663.1; JOINED.  
 DR EMBL; AF351815; AAK84663.1; JOINED.  
 DR EMBL; AF351816; AAK84663.1; JOINED.  
 DR EMBL; AF351817; AAK84663.1; JOINED.

DR EMBL; AF351818; AAK84663.1; JOINED.  
 DR EMBL; AF351819; AAK84663.1; JOINED.  
 DR EMBL; AF351820; AAK84663.1; JOINED.  
 DR EMBL; AF351821; AAK84663.1; JOINED.  
 DR EMBL; AF351822; AAK84663.1; JOINED.  
 DR EMBL; AF351823; AAK84663.1; JOINED.  
 DR Genew; HGNC:13887; ABCG8.  
 DR MIM; 605460; -.  
 DR MIM; 210250; -.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Glycoprotein; Transmembrane; Transport; Alternative splicing;  
 KW Polymorphism; Disease mutation.

FT	DOMAIN	1	416	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	417	437	1 (POTENTIAL).
FT	DOMAIN	438	447	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	448	468	2 (POTENTIAL).
FT	DOMAIN	469	492	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	493	513	3 (POTENTIAL).
FT	DOMAIN	514	531	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	532	552	4 (POTENTIAL).
FT	DOMAIN	553	569	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	570	590	5 (POTENTIAL).
FT	DOMAIN	591	639	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	640	660	6 (POTENTIAL).
FT	DOMAIN	661	673	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	619	619	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	376	376	Missing (in isoform 2).
FT				/FTId=VSP_000052.
FT	VARIANT	19	19	D -> H.
FT				/FTId=VAR_012250.
FT	VARIANT	54	54	Y -> C.
FT				/FTId=VAR_012251.
FT	VARIANT	184	184	R -> H (in sitosterolemia).
FT				/FTId=VAR_012252.
FT	VARIANT	231	231	P -> T (in sitosterolemia).
FT				/FTId=VAR_012253.
FT	VARIANT	238	238	E -> K.
FT				/FTId=VAR_012254.
FT	VARIANT	259	259	A -> V.
FT				/FTId=VAR_012255.
FT	VARIANT	263	263	R -> Q (in sitosterolemia).
FT				/FTId=VAR_012256.
FT	VARIANT	400	400	T -> K.
FT				/FTId=VAR_012257.
FT	VARIANT	405	405	R -> H (in sitosterolemia).
FT				/FTId=VAR_012258.
FT	VARIANT	501	501	L -> P (in sitosterolemia).
FT				/FTId=VAR_012259.
FT	VARIANT	543	543	R -> S (in sitosterolemia).
FT				/FTId=VAR_012260.
FT	VARIANT	570	570	Missing (in sitosterolemia).
FT				/FTId=VAR_012261.
FT	VARIANT	572	572	L -> P (in sitosterolemia).

```

FT                               /FTId=VAR_012262.
FT  VARIANT      574      574      G -> E (in sitosterolemia).
FT                               /FTId=VAR_012263.
FT  VARIANT      574      574      G -> R (in sitosterolemia).
FT                               /FTId=VAR_012264.
FT  VARIANT      575      575      G -> R.
FT                               /FTId=VAR_012265.
FT  VARIANT      596      596      L -> R (in sitosterolemia).
FT                               /FTId=VAR_012266.
FT  VARIANT      632      632      V -> A.
FT                               /FTId=VAR_012267.
SQ  SEQUENCE      673 AA;  75678 MW;  594AFD1D6C1BB50F CRC64;

```

```

Query Match      21.0%; Score 697; DB 1; Length 673;
Best Local Similarity  28.9%; Pred. No. 2.9e-42;
Matches 187; Conservative 124; Mismatches 241; Indels 96; Gaps 16;

```

```

Qy      8  TPGGSMGLQVNRGSQSSLEGAPAT-APEPHSLGILHASYSVSHRVR-PWWD-ITSCRQQW 64
      || : ||| | | | : : || : | | : || : : |
Db     16  TPQDTSGLQDRLFSSSDNSLYFTYSGQPNTLEVRDLNYQVDLASQVPWFELAQFKMPW 75

Qy     65  TRQI-----LKDVSLYVESGQIMCILGSSSGSKTTLDDAMSGRLGRAGTF-LGEVYV 115
      | : : : | | || : | || | : : || : | | | : : :
Db     76  TSPSCQNSCELGIQNLSFKVRSQMLAIIGSSGCGRASLLDVITGR-GHGGKIKSGQIWI 134

Qy    116  NGRALRREQFQDCFSYVLQSDTLLSSLTVRETLYHTALLAI-RRGNPGSFQKKVEAVMAE 174
      || : : : | : | | : || || || : | : : | : | : || |
Db    135  NGQPSSPQLVRKCVAVHRQHNLPLNTVRETLAFIAQMRLPRTFSQAQRDKRVEDVIAE 194

Qy    175  LSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLDPKVMLFDEPTTGLDCMTANQIVVL 234
      | | | : || : | : | || || || || || : | : : || : ||
Db    195  LRLRQCADTRVGNMYVRGLSGGERRRVSIGVQLLWNPGLILDEPTSGLDSTAHNLVKT 254

Qy    235  LVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEPH 294
      | || : || : : : || || : : || : : : | | : | : | : || :
Db    255  LSRLAKGNRLVLISLHQPRSDIFRLFDLVLLMTSGTPIYLGAAQHMVQYFTAIGYPCPRY 314

Qy    295  SNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHL----- 348
      ||| ||| : ||| : | : : : : | : : | : : : : : : : : :
Db    315  SNPADFYVDLTSIDRRSREQELATREKAQSLAALF-----LEKVRDLDDFLWK 362

Qy    349  -----KTLPM---VPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRL 390
      | : : | | | | : | : | | | : :
Db    363  AETKDLDEDTCESSVTPLDTNCLPSPTK-MPGAVQQFTTLIRQISNDFRDLPTLLIHG 421

Qy    391  LQNLIMGLFLLFFVLRVRSNVLKGAIQ----DRVGLLYQFVGATPYTGMLNAVNLFVPVLR 446
      : : | : : | : || | | | : : : : : : : : : : :
Db    422  AEACLMSMTIGFLYFG-----HGSIQLSFMDTAALLFMIGALIPFNVLVDVISKCYSER 475

Qy    447  AVSDQESQDGLYQKWQMMLAYALHVLFPFSVVATMIFSSVCYWTGLGHPEVARF----- 499
      | : | : || | | | | | : : | | | : |
Db    476  AMLYYELEDGLYTTGPYFFAKILGELPEHCAYII IYGMPYWLANLRPGLQPFLLHFLLV 535

Qy    500  -----GYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSVVALLSIAGVLVSGGFL 549
      : || | : | : | : : : : : : : : :
Db    536  WLNVFCCRIMALAAAALLPTFHMAFFS-----NALYNSFYLAG-----GFM 577

```

QY 550 RNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTN 597  
 | : : || : | : | | : : | : : | : : :  
 Db 578 INLSSLWTVPAWISKVSFLRWCFEGLMKIQFSRRTYKMPLGNLTIAVS 625

RESULT 6

ABG8\_RAT

ID ABG8\_RAT STANDARD; PRT; 694 AA.  
 AC P58428; Q8CIQ5; Q923R7;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 15-MAR-2004 (Rel. 43, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE ATP-binding cassette, sub-family G, member 8 (Sterolin-2).  
 GN ABCG8.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=Sprague-Dawley;  
 RX MEDLINE=21344600; PubMed=11452359;  
 RA Lu K., Lee M.-H., Hazard S., Brooks-Wilson A., Hidaka H., Kojima H.,  
 RA Ose L., Stalenhoef A.F.H., Mietinnen T., Bjorkhem I., Bruckert E.,  
 RA Pandya A., Brewer H.B. Jr., Salen G., Dean M., Srivastava A.K.,  
 RA Patel S.B.;  
 RT "Two genes that map to the STSL locus cause sitosterolemia: genomic  
 RT structure and spectrum of mutations involving sterolin-1 and  
 RT sterolin-2, encoded by ABCG5 and ABCG8, respectively."  
 RL Am. J. Hum. Genet. 69:278-290(2001).  
 RN [2]  
 RP REVISIONS TO 3-4.  
 RA Lu K., Yu H., Lee M.-H., Patel S.B.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND TISSUE SPECIFICITY.  
 RC STRAIN=GH, SHR, SHRSP, Sprague-Dawley, Wistar, Wistar Kyoto, and WKA;  
 RC TISSUE=Intestine, and Liver;  
 RX PubMed=12783625;  
 RA Yu H., Pandit B., Klett E., Lee M.-H., Lu K., Helou K., Ikeda I.,  
 RA Egashira N., Sato M., Klein R., Batta A., Salen G., Patel S.B.;  
 RT "The rat STSL locus: characterization, chromosomal assignment, and  
 RT genetic variations in sitosterolemic hypertensive rats."  
 RL BMC Cardiovasc. Disord. 3:4-4(2003).  
 CC -!- FUNCTION: Transporter that appears to play an indispensable role  
 CC in the selective transport of the dietary cholesterol in and out  
 CC of the enterocytes and in the selective sterol excretion by the  
 CC liver into bile.  
 CC -!- SUBUNIT: May form heterodimers with ABCG5 or be tightly coupled to  
 CC ABCG5 along a pathway regulating dietary-sterol absorption and  
 CC excretion (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=3;  
 CC IsoId=P58428-3; Sequence=Displayed;  
 CC Name=1;

```

CC      IsoId=P58428-1; Sequence=VSP_008767;
CC      Name=2;
CC      IsoId=P58428-2; Sequence=VSP_008767, VSP_000054;
CC      Note=No experimental confirmation available;
CC      -!- TISSUE SPECIFICITY: Highest expression in liver, with lower levels
CC          in small intestine and colon.
CC      -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)
CC          subfamily.
CC      -!- CAUTION: Seems to have a defective ATP-binding region.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF351785; AAK84831.2; -.
DR      EMBL; AY145899; AAN64276.1; -.
DR      EMBL; AF404109; AAK85393.1; -.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      Glycoprotein; Transmembrane; Transport; Alternative splicing.
FT      DOMAIN      1      434      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM      435      455      1 (POTENTIAL).
FT      DOMAIN      456      468      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM      469      489      2 (POTENTIAL).
FT      DOMAIN      490      517      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM      518      538      3 (POTENTIAL).
FT      DOMAIN      539      547      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM      548      568      4 (POTENTIAL).
FT      DOMAIN      569      590      CYTOPLASMIC (POTENTIAL).
FT      TRANSMEM      591      611      5 (POTENTIAL).
FT      DOMAIN      612      650      EXTRACELLULAR (POTENTIAL).
FT      TRANSMEM      651      671      6 (POTENTIAL).
FT      DOMAIN      672      694      CYTOPLASMIC (POTENTIAL).
FT      CARBOHYD      640      640      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      VARSPLIC      56      77      Missing (in isoform 1 and isoform 2).
FT      VARSPLIC      398      398      /FTId=VSP_008767.
FT      VARSPLIC      398      398      Missing (in isoform 2).
FT      VARSPLIC      398      398      /FTId=VSP_000054.
FT      CONFLICT      3      4      EK -> QT (IN REF. 3).
SQ      SEQUENCE      694 AA; 78236 MW; 67F67C195F417587 CRC64;

```

```

Query Match      20.8%; Score 692.5; DB 1; Length 694;
Best Local Similarity 29.3%; Pred. No. 6.4e-42;
Matches 189; Conservative 122; Mismatches 255; Indels 79; Gaps 17;

```

```

Qy      34 EPH-SLGILHASYSVSHRVRPW-----WDITSCRQQWTRQILKDVSLYVESGQIM 82
        :|| ||| | | :: :| ||      |      : | | ::| | ||::
Db      67 DPHMSLG-LSESVDMAQV-PWFEQLAQFKLPWRSRGSQDSWDLGI-RNLSFKVRSGQML 123

```





RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=99065313; PubMed=9850061;  
 RA Allikmets R., Schriml L.M., Hutchinson A., Romano-Spica V., Dean M.;  
 RT "A human placenta-specific ATP-binding cassette gene (ABCP) on  
 RT chromosome 4q22 that is involved in multidrug resistance.";  
 RL Cancer Res. 58:5337-5339(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast cancer;  
 RX MEDLINE=99080071; PubMed=9861027;  
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
 RA Ross D.D.;  
 RT "A multidrug resistance transporter from human MCF-7 breast cancer  
 RT cells.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:15665-15670(1998).  
 RN [3]  
 RP ERRATUM.  
 RA Doyle L.A., Yang W., Abruzzo L.V., Krogmann T., Gao Y., Rishi A.K.,  
 RA Ross D.D.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:2569-2569(1999).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Kage K., Tsukahara S., Sugiyama T., Asada S., Ishikawa E., Tsuruo T.,  
 RA Sugimoto Y.;  
 RT "Breast cancer resistance protein constitutes a 140-kDa complex as a  
 RT homodimer.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE OF 198-655 FROM N.A.  
 RC TISSUE=Placenta;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,  
 RA Ninomiya K., Iwayanagi T.;  
 RT "NEDO human cDNA sequencing project.";  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC -!- FUNCTION: Xenobiotic transporter that appears to play a major role  
 CC in the multidrug resistance phenotype of a specific MCF-7 breast  
 CC cancer cell line. When overexpressed, the transfected cells become  
 CC resistant to mitoxantrone, daunorubicin and doxorubicin, display  
 CC diminished intracellular accumulation of daunorubicin, and  
 CC manifest an ATP-dependent increase in the efflux of rhodamine 123.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

DR EMBL; AF103796; AAD09188.1; -.
DR EMBL; AF098951; AAC97367.1; -.
DR EMBL; AB056867; BAB39212.1; -.
DR EMBL; AK002040; BAA92050.1; -.
DR Genew; HGNC:74; ABCG2.
DR MIM; 603756; -.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0005524; F:ATP binding; TAS.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. . .; TAS.
DR GO; GO:0005215; F:transporter activity; TAS.
DR GO; GO:0008559; F:xenobiotic-transporting ATPase activity; TAS.
DR GO; GO:0009315; P:drug resistance; TAS.
DR GO; GO:0006810; P:transport; TAS.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; FALSE_NEG.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transmembrane; Transport.
FT DOMAIN 1 395 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 396 416 POTENTIAL.
FT DOMAIN 417 428 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 429 449 POTENTIAL.
FT DOMAIN 450 477 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 478 498 POTENTIAL.
FT DOMAIN 499 506 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 507 527 POTENTIAL.
FT DOMAIN 528 535 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 536 556 POTENTIAL.
FT DOMAIN 557 630 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 631 651 POTENTIAL.
FT DOMAIN 652 655 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 80 87 ATP (POTENTIAL).
FT CARBOHYD 418 418 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 24 24 V -> A (IN REF. 2 AND 4).
FT CONFLICT 166 166 E -> Q (IN REF. 2 AND 4).
FT CONFLICT 208 208 F -> S (IN REF. 1).
FT CONFLICT 315 316 MISSING (IN REF. 5).
FT CONFLICT 482 482 R -> T (IN REF. 2).
SQ SEQUENCE 655 AA; 72343 MW; 89A6D3511DC5CCE0 CRC64;

```

Query Match 20.3%; Score 676.5; DB 1; Length 655;  
 Best Local Similarity 29.0%; Pred. No. 8.2e-41;  
 Matches 181; Conservative 137; Mismatches 251; Indels 55; Gaps 18;

QY 21 SQSSLEGAPATAP---EPHSLGILHASYSVSHRVRPWWDITSCRQQWTRQILKDVSLYVE 77

	:       : :   : : :: :    :   : :   :: :	
Db	13 SQGNTNGFPATVSNDLKAFTEGAVLSFHNICYRVKLKSGFLPCRKPVEKEILSNINGIMK	72
Qy	78 SGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDT	137
	:     :    ::    ::   :  :  :       :  :  :	
Db	73 PG-LNAILGPTGGGKSSLLDVLAAARKDPSG-LSGDVLING-APRPANFKCNSGYVQDDV	129
Qy	138 LLSSLTVRETLHYTALLAIRRGNPG-SFQKKVEAVMAELSLSHVADRLLIGNYSLGGISTG	196
	:: :      :   :  : :  :        :  :  :	
Db	130 VMGTLTVRENLQFSAALRLATTMTNHEKNERINRVIEELGLDKVADSKVGTQFIRGVSGG	189
Qy	197 ERRRVSIAAQLLQDPKVMFLDEPTTGLDCMTANQIVVLLVELARRNRIVVLTIHQPRSEL	256
	:     : :    :            ::   ::::   : :      :	
Db	190 ERKRTSIGMELITDPSILFLDEPTTGLDSSTANAVLLLLKRMSKQGRTIIFSIIHQPRYSI	249
Qy	257 FQLFDKIAILSFGELIFCGTPAEMLDFFNDGCGYPCPEHSNPFDFYMDLTSVDTQ----	312
	:    : : :    :      :       :     :: : :  : ::	
Db	250 FKLFDLTLTLLASGRLMFHGPAQEALGYFESAGYHCEAYNNPADFFLDIINGDSTAVALNR	309
Qy	313 ERE-----IETSKR----VQMIESAYKKSACHKT-----LKNIERMKHLKTLPMVPF	356
	:        : : :     : :      :  :  : : :	
Db	310 EEDFKATEIIEPSKQDKPLIEKLAEIYVNSSFYKETKAELHQLSGGEKKKKITVFKEISY	369
Qy	357 KTKDSPGVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFL--LFFVLRVRSNVLKG	414
	:  : :  :  :       :: :    : :    :	
Db	370 TT----SFCHQLRWVSKRSEFKNLLGNPQASIAQIIIVTVVLGLVIGAIYFGLKNDST----	421
Qy	415 AIQDRVGLLYQFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQMMLAYAL-HVLP	473
	:   : :  : :        : :     :      :	
Db	422 GIQNRAGVLF-FLTTNQCFSSVSAVELFVVEKKLFIHEYISGYRVSYYFLGKLLSDLLP	480
Qy	474 FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNIVNSV	533
	:: :  : :  :       :   : : : :    :  :	
Db	481 MRMLPSIIFTCIVYFMLGLKPKADAFFVMMFTLM---MVAYSASSMALAIAAGQSVVSV	537
Qy	534 VALLSIAGV--LVGSGFLRNIQEMPIPKIISYFTFQKYCSEILVNNEFYGLNFTCGSSN	591
	::    : :     : : :   : :	
Db	538 TLLMTICFVFMFIFSGLLVNLTTIASWLSWLQYFSIPRYGFTALQHNEFLGQNFPCPG---	594
Qy	592 VSVTTNPMCAFTQGIQFIEKTCPG	615
	::       :	
Db	595 LNATGNNPCNYA-----TCTG	610

# RESULT 8

YOH5\_YEAST

ID YOH5\_YEAST STANDARD; PRT; 1294 AA.

AC Q08234; Q08233;

DT 01-NOV-1997 (Rel. 35, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Probable ATP-dependent transporter YOL074C/YOL075C.

GN YOL074C/YOL075C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97321807; PubMed=9178509;
RA      Tzermia M., Katsoulou C., Alexandraki D.;
RT      "Sequence analysis of a 33.2 kb segment from the left arm of yeast
RT      chromosome XV reveals eight known genes and ten new open reading
RT      frames including homologues of ABC transporters, inositol
RT      phosphatases and human expressed sequence tags.";
RL      Yeast 13:583-589(1997).
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC      -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; Z74817; CAA99085.1; -.
DR      EMBL; Z74816; CAA99084.1; -.
DR      PIR; S77690; S77690.
DR      GermOnline; 143497; -.
DR      SGD; S0005435; YOL075C.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      Pfam; PF00005; ABC_tran; 2.
DR      ProDom; PD000006; ABC_transporter; 2.
DR      SMART; SM00382; AAA; 2.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW      Hypothetical protein; ATP-binding; Transmembrane; Glycoprotein;
KW      Transport; Repeat.
FT      TRANSMEM      376      396      POTENTIAL.
FT      TRANSMEM      496      516      POTENTIAL.
FT      TRANSMEM      531      551      POTENTIAL.
FT      TRANSMEM      605      625      POTENTIAL.
FT      TRANSMEM      1039     1059     POTENTIAL.
FT      TRANSMEM      1121     1141     POTENTIAL.
FT      TRANSMEM      1267     1287     POTENTIAL.
FT      NP_BIND        62       69      ATP (POTENTIAL).
FT      NP_BIND       727      734      ATP (POTENTIAL).
FT      CARBOHYD        41       41      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD        86       86      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD       101      101      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD       151      151      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD       341      341      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD       349      349      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD       371      371      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD       528      528      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD       983      983      N-LINKED (GLCNAC. . .) (POTENTIAL).
FT      CARBOHYD      1062     1062      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ      SEQUENCE      1294 AA;  145157 MW;  C555500A45E9284E CRC64;

Query Match          18.9%;  Score 627;  DB 1;  Length 1294;
Best Local Similarity 31.7%;  Pred. No. 6.9e-37;

```

Matches 181; Conservative 106; Mismatches 228; Indels 56; Gaps 19;

```
Qy      65 TRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLL-----GEVYVNGRA 119
      |::||: |:  : | | | :| | ||||:|::|: :||| |:: |  | : |
Db      706 TKEILQSVNAIFKPGMINAIMGPSGSGKSSLLNLISGRL-KSSVFAKFDTSGSIMENDIQ 764

Qy      120 LRREQFQDCFSYVLQSDT-LLSSLTVRETLHYTALLAIRRGNPGSFQKKVEAVMAELSLS 178
      :  |:: ||| | | ||:|::| | | | :  : : : : | |
Db      765 VSELMFKNVCSYVSQDDHLLAALT VKETLKYAAALRLHHLTEAERMERTDNLIRSLGLK 824

Qy      179 HVADRLIGNYSLGGISTGERRRRVSIAAQLQDPKVMLEFDEPTTGLDCMTANQIVVLLVEL 238
      | : :||| : ||| ||:|::| | | | :| | | | | | | | : | : : | : |
Db      825 HCENNIIGNEFVKGISGGEKRRVTMGVQLLNDPPILLLDEPTSGLDSTSATILEILEKL 884

Qy      239 AR-RNRIVVLTIHQPRSELFQLFDKIAILS-FGELIFCGTPAEMLDFFNDGCGYPCPEHSN 296
      | : : :::| | | | | | | | : | : : | | | | | : | : | | | : |
Db      885 CREQGKTIIITIHQPRSELFKRFGNVLLAKSGRTAFNGSPDEMIAYFTELGYNCPSFTN 944

Qy      297 PFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPF 356
      ||::| | ||:|::| | | : | | : | | | | | | | : : | | | :
Db      945 VADFFLDLISVNTQNEQNEISSRARVEKILSAWK-----ANMDN-ESLSPTPISEK 994

Qy      357 KTKDSPGVFSKLGVLRRVTRNLV-----RNKLAVITRLLQNLIMGLFLL 401
      :  |::  : | : | | | | | | | | | | : : : | : | : | :
Db      995 QQYSQESFFTEYSEFVRK-PANLVLAYIVNVKRQFTTTTRRSFDSLARIAQIPGLGVIFA 1053

Qy      402 FFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFPVLRVSDQESQDGLYQKW 461
      | | : | : | : | : | : | | : : | | : | : | : | : |
Db      1054 LFFAPVKHNYT--SISNRLGLAQEST-ALYFVGMLGNLACYPTERDYFEEYNDNVYGIA 1110

Qy      462 QMMLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVARFGYFSAALLAPHLI---GEFLTL 518
      ||| || | | : : : : : || | | | | : : : || | :
Db      1111 PFFLAYMTLELPLSALASVLYAVFTVLACGL-PRTA--GNFFATVYCSFIVTCCGERLGI 1167

Qy      519 VLLGIVQNPN-IVNSVVALLSIAGVLVSGSGLRNIQEMPIPFKIISYFTFQKYCSEILVV 577
      :  : | : | : : | | : | | | | | : | | | : | : | :
Db      1168 MTNTFFFERPGFVVNCISIILSIGTQMSGLMSL---GMSRVLKGFNYLNPVGYTSMIIN 1223

Qy      578 NEFYG-LNFTC--GSSNVSVTTNPMCAFTQG 605
      | | | | | | | | | | | |
Db      1224 FAFPGNLKLTCEDGGKNSDGT----CEFANG 1250
```

RESULT 9

WHIT\_LUCCU

ID WHIT\_LUCCU STANDARD; PRT; 677 AA.

AC Q05360;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE White protein.

GN W.

OS *Lucilia cuprina* (Greenbottle fly) (Australian sheep blowfly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Oestroidea;

OC Calliphoridae; *Lucilia*.

OX NCBI\_TaxID=7375;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97087158; PubMed=8933176;
RA      Garcia R.L., Perkins H.D., Howells A.J.;
RT      "The structure, sequence and developmental pattern of expression of
RT      the white gene in the blowfly Lucilia cuprina.";
RL      Insect Mol. Biol. 5:251-260(1996).
RN      [2]
RP      SEQUENCE OF 490-584 FROM N.A.
RX      MEDLINE=90264941; PubMed=1971656;
RA      Elizur A., Vacek A.T., Howells A.J.;
RT      "Cloning and characterization of the white and topaz eye color genes
RT      from the sheep blowfly Lucilia cuprina.";
RL      J. Mol. Evol. 30:347-358(1990).
CC      -!- FUNCTION: May be part of a membrane-spanning permease system
CC      necessary for the transport of pigment precursors into pigment
CC      cells responsible for eye color.
CC      -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC      -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U38899; AAA82057.1; -.
DR      EMBL; X53265; CAA37365.1; -.
DR      InterPro; IPR003593; AAA_ATPase.
DR      InterPro; IPR003439; ABC_transporter.
DR      InterPro; IPR005284; Pigment_permease.
DR      Pfam; PF00005; ABC_tran; 1.
DR      ProDom; PD000006; ABC_transporter; 1.
DR      SMART; SM00382; AAA; 1.
DR      TIGRFAMs; TIGR00955; 3a01204; 1.
DR      PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR      PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
KW      Pigment; ATP-binding; Transmembrane; Transport.
FT      NP_BIND      119      126      ATP (POTENTIAL).
FT      TRANSMEM     431      451      POTENTIAL.
FT      TRANSMEM     456      476      POTENTIAL.
FT      TRANSMEM     506      526      POTENTIAL.
FT      TRANSMEM     534      554      POTENTIAL.
FT      TRANSMEM     563      583      POTENTIAL.
FT      TRANSMEM     647      667      POTENTIAL.
SQ      SEQUENCE     677 AA;  75365 MW;  D16FC11C97EED51D CRC64;

```

```

Query Match          18.7%; Score 623; DB 1; Length 677;
Best Local Similarity 27.0%; Pred. No. 5.8e-37;
Matches 188; Conservative 144; Mismatches 260; Indels 104; Gaps 19;

```

```

Qy      8 TPGGSMGLQVNRGSQSSLEGAPATAPEPHSLGILHASYSVSHRVRPWW-----DITSC 60
      |||      :|| :   :   ||| :: || :: |      ::
Db      27 TPG-----TLEASAINSGFSKSYGSLVSNESASEKLTYSWCNLDVFGVHQP 73

```





OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92160395; PubMed=1789009;  
 RA Purnelle B., Skala J., Goffeau A.;  
 RT "The product of the YCR105 gene located on the chromosome III from  
 RT *Saccharomyces cerevisiae* presents homologies to ATP-dependent  
 RT permeases.";  
 RL Yeast 7:867-872(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92327849; PubMed=1626432;  
 RA Skala J., Purnelle B., Goffeau A.;  
 RT "The complete sequence of a 10.8 kb segment distal of *SUF2* on the  
 RT right arm of chromosome III from *Saccharomyces cerevisiae* reveals  
 RT seven open reading frames including the *RVS161*, *ADP1* and *PGK* genes.";  
 RL Yeast 8:409-417(1992).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X59720; CAA42328.1; -.  
 DR PIR; S19421; S19421.  
 DR GermOnline; 138916; -.  
 DR SGD; S0000604; ADP1.  
 DR GO; GO:0005783; C:endoplasmic reticulum; IDA.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Transmembrane; Glycoprotein; Transport; Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1049 PROBABLE ATP-DEPENDENT PERMEASE.  
 FT NP\_BIND 423 430 ATP (BY SIMILARITY).  
 FT TRANSMEM 325 345 POTENTIAL.  
 FT TRANSMEM 464 481 POTENTIAL.  
 FT TRANSMEM 794 814 POTENTIAL.  
 FT TRANSMEM 829 849 POTENTIAL.  
 FT TRANSMEM 878 898 POTENTIAL.  
 FT TRANSMEM 910 930 POTENTIAL.  
 FT TRANSMEM 938 958 POTENTIAL.  
 FT TRANSMEM 1001 1021 POTENTIAL.  
 FT TRANSMEM 1025 1045 POTENTIAL.  
 FT CARBOHYD 50 50 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 221 221 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 815 815 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 960 960 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 971 971 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1049 AA; 117231 MW; ABC9CE54BCFDF6A3 CRC64;

Query Match 18.7%; Score 621; DB 1; Length 1049;  
 Best Local Similarity 28.6%; Pred. No. 1.4e-36;  
 Matches 196; Conservative 111; Mismatches 223; Indels 156; Gaps 22;

Qy 68 ILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRLGRAGTFLGEVYVNGRALRREQFQD 127  
 :| ::| |: |||: |:| ||:||||||| :: : :| | : ||| :: |: |  
 Db 405 VLNEISGIVKPGQILAIMGGSGAGKTTLLDILAMK-RKTGHVSGSIKVNIGISMDRKSFSK 463

Qy 128 CFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQKK---VEAVMAELSLSHVADRL 184  
 :| | | || :||| ||: :||| : : ||: | | |: | | : : ||:  
 Db 464 IIGFVDQDDFLPTLTVFETVLNSALLRLPKAL--SFEAKKARVYKVLEELRIIDIKDRI 521

Qy 185 IGNYSLGGISTGERRRVSIQAQLLDQPKVMLFDEPTTGLDCMTANQIVVLLVELAR-RNR 243  
 ||| ||| ||:||||| |: | |: |||:||| || :: || |: ||  
 Db 522 IGNEFDRGISGGEKRRVSIACELVTSPLVLFLDEPTSGLDASNANNVIECLVRLSSDYNR 581

Qy 244 IVVLTIHQPRSELFQFLDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPFDFYMD 303  
 :||:||||| :| ||||: :|| ||::: | :: :| : || ||:: | |: :|  
 Db 582 TLVLSIHQPRSNIFYLFDKLVLLSKGEMVYSGNAKKVSEFLRNEGYICPDNYNIADYLID 641

Qy 304 LT-SVDTQSKEREI----- 316  
 :| | | |  
 Db 642 ITFEAGPQGKRRRIRNISDLEAGTDTNDIDNTIHQTTFTSSDGTQREWAHLAAHRDEIR 701

Qy 317 -----ETSKRVQMIESAYKKSACHKTLKNIERM----- 345  
 | :: || | : : || :  
 Db 702 SLLRDEEDVEGTDGRRGATEIDLNTKLLHDKYKDSVYYAELSQEIEEVLSEGDEESNVLN 761

Qy 346 KHLKTLPMVPFKTKDSPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFLLFFVL 405  
 | | : | | :| :| | :|: || :: | ::| ||  
 Db 762 GDLPT-----GQQSAGFLQQLSILNSRSFKNMYRNPKLLLGNYLTLILLSLFLGTLTY 814

Qy 406 RVRSNVLKGAIQDRVGLLY---QFVGATPYTGMLNAVNLFVPLRAVSDQESQDGLYQKWQ 462  
 | || : | |:|:| : : | :||: : | : | : : | : |  
 Db 815 NV-SNDISG-FQNRMGLEFFILTIFYGFVTFGL----SSFALERIIFIKERSNNYYSP-- 866

Qy 463 MMLAYAL-----HVLPFSSVATMIFSSVCYWTGLGHPEVARFGYFSAALLAPHLIGEFLT 517  
 ||| : |:| || :: | : | ||: : | : :| : | |  
 Db 867 --LAYYISKIMSEVVPLRVPPILLSLIVYPMTGLNMKDNAF-FKCIGILILFNLGISLE 923

Qy 518 LVLLGIV---QNPNIIVNSVALLSIAGVLVGSGFLRNIQEMP-IPFKIISYFTFQKYCSE 573  
 :: :||: | :|: ||: || | | : || | : : : ||: | | |  
 Db 924 ILTIGIIFEDLNNSSIILSVLVLL---GSLLFSGLFINTKNITNVAFKYLKNFSVFYYAYE 980

Qy 574 ILVVNEF-----YGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRETMNF 624  
 |:| | |||| | || | |  
 Db 981 SLLINEVKTLMLKERKYGLNI-----EVPGATILSTFGF 1014

Qy 625 LILYSFIPALVILGI--VVFKIRDHL 648  
 :: : : : || : ||| | :|  
 Db 1015 -VVQNLVFDIKILALFNVVFLIMGYL 1039

RESULT 11

WHIT\_ANOGA

ID WHIT ANOGA STANDARD; PRT; 695 AA.  
AC Q27256; Q17006;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE White protein.  
GN W.  
OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
OX NCBI\_TaxID=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Suakoko / G3;  
RX MEDLINE=96423158; PubMed=8825759;  
RA Besansky N.J., Bedell J.A., Benedict M.Q., Mukabayire O., Hilfiker D.,  
RA Collins F.H.;  
RT "Cloning and characterization of the white gene from Anopheles  
RT gambiae.";  
RL Insect Mol. Biol. 4:217-231(1995).  
CC -!- FUNCTION: May be part of a membrane-spanning permease system  
CC necessary for the transport of pigment precursors into pigment  
CC cells responsible for eye color.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U29486; AAC46995.1; -.  
DR EMBL; U29485; AAC46994.1; -.  
DR EMBL; U29484; AAC47423.1; -.  
DR InterPro; IPR003593; AAA\_ATPase.  
DR InterPro; IPR003439; ABC\_transporter.  
DR InterPro; IPR008965; Cellul\_bind.  
DR InterPro; IPR005284; Pigment\_permease.  
DR Pfam; PF00005; ABC\_tran; 1.  
DR ProDom; PD000006; ABC\_transporter; 1.  
DR SMART; SM00382; AAA; 1.  
DR TIGRFAMs; TIGR00955; 3a01204; 1.  
DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
KW Pigment; ATP-binding; Transmembrane; Transport.  
FT NP\_BIND 133 140 ATP (POTENTIAL).  
FT NP\_BIND 288 295 ATP (POTENTIAL).  
FT TRANSMEM 444 464 POTENTIAL.  
FT TRANSMEM 474 494 POTENTIAL.  
FT TRANSMEM 524 544 POTENTIAL.

FT	TRANSMEM	552	572	POTENTIAL.
FT	TRANSMEM	581	601	POTENTIAL.
FT	TRANSMEM	669	689	POTENTIAL.
FT	CARBOHYD	472	472	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	645	645	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	100	100	N -> S (IN REF. 1; AAC47423).
FT	CONFLICT	691	693	SRS -> YAR (IN REF. 1; AAC47423).
SQ	SEQUENCE	695 AA;	77218 MW;	EE8B9517239B2961 CRC64;

Query Match 18.3%; Score 607.5; DB 1; Length 695;  
 Best Local Similarity 28.4%; Pred. No. 7.7e-36;  
 Matches 170; Conservative 124; Mismatches 208; Indels 97; Gaps 17;

Qy	58	TSCRQQWTRQ-----ILKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGT	108
		:   :   :   :   :   :   :   :   :	
Db	96	TRLRNCCTRQRKDFNPRKHLKKNVTGVAKSGELLAVMGSSGAGKTTLLNALAFR-SPPGV	154
Qy	109	FLGEVYV---NGRALRREQFQDCFSYVLQSDTLSSLTVRETLHYTALLAIRGNPGSF-	164
		:       :   :   :   :   :   :   :   :   :	
Db	155	KISPNAVRALNGVPVNAEQLRARCAYVQDDLFIPSLTTREHLLFQAMLRMGRDVPASVK	214
Qy	165	QKKVEAVMAELSLSHVADRLIGNYS-LGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGL	223
		:   :   :   :   :   :   :   :   :   :   :	
Db	215	QHRVQEVQLQELSLVKCADTIIGAPGRIKGLSGGERKRLAFASETLTDPHLLLCDEPTSGL	274
Qy	224	DCMTANQIVVLLVELARRNRIVVLTIIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDF	283
		: : :   :   : : :   :   :   :   :   :	
Db	275	DSFMAHSVQLQVLKGMAMKGKTIILTIHQPSSELYCLFDKILLVAEGRVAFLGSPYQSAEF	334
Qy	284	FNDCGYPCPEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIE	343
		:       :       : : : : :   :   :   :   :	
Db	335	FSQLGIPCPPNYPADFYVQMLAI---APAKEAECDRMIKKICDSFAVSPiAREVLETAS	391
Qy	344	RMKHLKTLPMVPFKTKDSPGVFSKL-GV-----LLRRVTRNLVNRNKLAVI	387
		: :     :     : : : :	
Db	392	-----VAGKGMDEPYMLQQVEGVGSTGYRSSWWTQFYCILWRSWLSVLKDPMLVK	441
Qy	388	TRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRVGL-----LYQFVGATPYTGMNLAVNL	441
		: : : :   :   :   :   :   : : :	
Db	442	VRLQTAMVATLI-----GSIYFGQVLDQDGMNINGSFLFLTNMTFQNVFAVINV	493
Qy	442	FPVLRVSDQESQDGLYQKWQMMLAYALHVLPPFSVVATMIFSSVCYWTGLGHPEVARFGY	501
		:   :   :   :   :   :   :   :   :	
Db	494	FSAELPVFLREKRSRLRYVDTYFLGKTIAELPLFIAPFVFTSITYPMIGL-----RTG-	547
Qy	502	FSAALLAPHLIGEFLTLVLLGIVQNPN-----IVNSVVALLSIA-----GVLVGSG	547
		:   :   :   : : : : : : :   :	
Db	548	-----ATHYL---TTLFIVTLVANVSTSGYLISCASSSISMALSVGPPVVIPFLIFGG	598
Qy	548	FLRNIQEMPIPFKIISYFTFQKYCSEILVVNEFYGL-----NFTCGSSNVSVTT	596
		:     :   : :   :   : : : :       : :	
Db	599	FFLNSASVPAYFKYLSYLSWFRYANEALLINQWSTVVDGEIACTRANVTCPRSEIILET	657

RESULT 12  
 WHIT\_DROME  
 ID WHIT\_DROME STANDARD; PRT; 687 AA.

AC P10090; Q9V3A2; Q9XY33;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE White protein.  
 GN W OR EG:BACN33B1.1 OR CG2759.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Head;  
 RX MEDLINE=90221897; PubMed=2109311;  
 RA Pepling M., Mount S.M.;  
 RT "Sequence of a cDNA from the *Drosophila melanogaster* white gene.";  
 RL Nucleic Acids Res. 18:1633-1633(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85134865; PubMed=6084717;  
 RA O'Hare K., Murphy C., Levis R., Rubin G.M.;  
 RT "DNA sequence of the white locus of *Drosophila melanogaster*.";  
 RL J. Mol. Biol. 180:437-455(1984).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21100348; PubMed=11156992;  
 RA Lukacsovich T., Asztalos Z., Awano W., Baba K., Kondo S., Niwa S.,  
 RA Yamamoto D.;  
 RT "Dual-tagging gene trap of novel genes in *Drosophila melanogaster*.";  
 RL Genetics 157:727-742(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Oregon-R;  
 RX MEDLINE=20196011; PubMed=10731137;  
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,  
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,  
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,  
 RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,  
 RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,  
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourkioti F.,  
 RA Beinert N., Dowe G., Schaefer U., Jaeckle H., Bucheton A.,  
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,  
 RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,  
 RA Glover D.M.;  
 RT "From sequence to chromosome: the tip of the X chromosome of *D.*  
 RT *melanogaster*.";  
 RL Science 287:2220-2222(2000).  
 RN [6]  
 RP SEQUENCE OF 224-331 FROM N.A.  
 RX MEDLINE=89339145; PubMed=2503416;  
 RA Tearle R.G., Belote J.M., McKeown M., Baker B.S., Howells A.J.;  
 RT "Cloning and characterization of the scarlet gene of *Drosophila*  
 RT *melanogaster*.";  
 RL Genetics 122:595-606(1989).  
 CC -!- FUNCTION: Part of a membrane-spanning permease system necessary  
 CC for the transport of pigment precursors into pigment cells  
 CC responsible for eye color. White dimerize with brown for the  
 CC transport of guanine and with scarlet for the transport of  
 CC tryptophan.  
 CC -!- SUBUNIT: Heterodimer of white with either brown or scarlet.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X51749; CAA36038.1; -.  
 DR EMBL; X02974; CAA26716.1; -.  
 DR EMBL; AB028139; BAA78210.1; -.  
 DR EMBL; AE003425; AAF45826.1; -.  
 DR EMBL; AL133506; CAB65847.1; -.  
 DR EMBL; X76202; CAA53795.1; -.  
 DR PIR; S08635; FYFFW.  
 DR FlyBase; FBgn0003996; w.  
 DR GO; GO:0004888; F:transmembrane receptor activity; NAS.  
 DR GO; GO:0006727; P:ommochrome biosynthesis; IMP.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00955; 3a01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP\_BIND 130 137 ATP (BY SIMILARITY).  
 FT TRANSMEM 435 453 POTENTIAL.  
 FT TRANSMEM 465 485 POTENTIAL.  
 FT TRANSMEM 515 533 POTENTIAL.  
 FT TRANSMEM 542 563 POTENTIAL.  
 FT TRANSMEM 576 594 POTENTIAL.  
 FT TRANSMEM 659 678 POTENTIAL.  
 FT CONFLICT 25 29 GDSGA -> LIFEIPYHCRVTAD (IN REF. 2 AND  
 FT 3).  
 FT CONFLICT 49 49 L -> R (IN REF. 4 AND 5).  
 FT CONFLICT 335 371 VGAQCPTNYPADFYVQVLAVVPGREIESRDRIAKIC ->  
 FT ITLHLNSYPAWVPSVLPTTIRRTFTYRCWPLCPDGRSSPVI  
 FT GSPRYG (IN REF. 3).  
 SQ SEQUENCE 687 AA; 75672 MW; 24AFAD799DE0D396 CRC64;

Query Match 18.1%; Score 602.5; DB 1; Length 687;  
 Best Local Similarity 28.8%; Pred. No. 1.7e-35;  
 Matches 180; Conservative 131; Mismatches 220; Indels 95; Gaps 19;

Qy 66 RQILKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGR--LGRAGTFLGEVYVNGRALRRE 123  
 : :||:| |::: :|||:| ||||:|:: | | : | :||: : :  
 Db 110 KHLKNCVCGVAYPGELLAVMGSSGAGKTTLLNALAFRSPQGIQVSPSGMRLLNGQPVDK 169  
 Qy 124 QFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRGNPGSFQK---KVEAVMAELSLSHV 180  
 : | :|| | | : ||| || | : | :| :||: :|: | :|||  
 Db 170 EMQARCAVYVQDDLFIGSLTAREHLIFQAM--VRMPRHLYRQRVARVDQVIQELSLSKC 227  
 Qy 181 ADRLIG-NYSLGGISTGERRRVSIAAQLLQDPKVMFLDEPTTGLDCMTANQIVVLLVELA 239  
 :|| : |:| |||:|:: |:: | || :||: ||| ||: :| :| :|:  
 Db 228 QHTIIGVPGRVKGLSGGERKRLAFASEALTDPLLICDEPTSGLDSTAHVSVVQLKKLS 287  
 Qy 240 RRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDGYPCEHSNPF 299  
 : : : |:||||| ||||:||||| :|| : | :| |||:| :||: | || : || |  
 Db 288 QKGKTVILTIHQPSSELFELFDKILLMAEGRVAFGLTPSEAVDFFSYVGAQCPTNYPAD 347  
 Qy 300 FYMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTK 359  
 ||: : :| ||||: |: | : | | :||:| | |

```

Db      348 FYVQVLAV---VPGREIESRDRIAKICDNFAIS----KVARDMEQLLATKNLE----KPL 396
Qy      360 DSP-----GVFSKLGVLRRVTRNLVRNKLAVITRLLQNLIMGLFL-LFFVLRVRSN 410
      : |           | :   :| |   ::: | |   ||:|   :: : : | |: : :
Db      397 EQPENGYTYKATWFMQFRAVLWRSWLSVLKEPLLVKVRLIQTTMVAILIGLIFLGQQLTQ 456
Qy      411 VLKGAIQDRVGLLYQFVGATPYTGMLNAVNLFVPVLRVAVSDQESQDGLYQKWOMMLAYALH 470
      |   : :   | :: | :   : :   :|:|   |   :|:: ||:   |   :
Db      457 V---GVMNINGAIFLFLTNTMTQNVFATINVFTSELPVFMREARSRLRYCDTYFLGKTIA 513
Qy      471 VLPFSVAVATMIFSSVCYWTGLHPEVARFGYFSAALLAPHLIGEFLLTVLLGIVQNPNIIV 530
      ||   :   ::|::: |   :||   |   |   |   ||: :|   |:
Db      514 ELPLFLTVPVLFVTAIAYPMIGLRAGVLHF-----FNCLALVTLV--ANVS 556
Qy      531 NSVVALLSIAG-----VLVSGSGLRNIQEMPIPFKIISYFTFQKYCSEI 574
      |   |:| |           |: | |   |: |   :| | : : | :|
Db      557 TSFGYLISCASSSTSMALSVGPPVIIPFLLFGGFFLNSGSVPVYLKWLSYLSWFRYANEG 616
Qy      575 LVVNEFYGL---NFTCGSSNVSVTTNPMCAFTQGIQFIEKTCF--GATSRFTMFLILYS 629
      |:|::: :   :| |||           ||| |   |:| |   :
Db      617 LLINQWADVEPGEISCTSSNT-----TCPSSGKVIETLNFS--A 655
Qy      630 FIP---ALVILGIVVFKIRDHLISR 651
      :|   | || || |: : :| |
Db      656 DLPLDYVGLAIL-IVSFRVLAYLALR 680

```

ABG1 HUMAN

```

ID      ABG1 HUMAN                STANDARD;          PRT;      678 AA.
AC      P45844; Q9BXK6; Q9BXK7; Q9BXK8; Q9BXK9; Q9BXL0; Q9BXL1; Q9BXL2;
AC      Q9BXL3; Q9BXL4;
DT      01-NOV-1995 (Rel. 32, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      ATP-binding cassette, sub-family G, member 1 (White protein homolog)
DE      (ATP-binding cassette transporter 8).
GN      ABCG1 OR ABC8 OR WHT1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE OF 3-678 FROM N.A. (ISOFORMS 1 AND 4).
RC      TISSUE=Retina;
RX      MEDLINE=96256850; PubMed=8659545;
RA      Chen H.M., Rossier C., Lalioti M.D., Lynn A., Chakravarti A.,
RA      Perrin G., Antonarakis S.E.;
RT      "Cloning of the cDNA for a human homologue of the Drosophila white
RT      gene and mapping to chromosome 21q22.3.";
RL      Am. J. Hum. Genet. 59:66-75(1996).
RN      [2]
RP      SEQUENCE FROM N.A. (ISOFORM 1).
RX      MEDLINE=20289799; PubMed=10830953;
RA      Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA      Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA      Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.

```

RX MEDLINE=20289799; PubMed=10830953;

RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,

RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.

RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,



RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,  
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,  
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,  
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;  
 RT "The DNA sequence of human chromosome 21.";  
 RL Nature 405:311-319(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=20408883; PubMed=10950923;  
 RA Berry A., Scott H.S., Kudoh J., Talior I., Korostishevsky M.,  
 RA Wattenhofer M., Guipponi M., Barras C., Rossier C., Shibuya K.,  
 RA Wang J., Kawasaki K., Asakawa S., Minoshima S., Shimizu N.,  
 RA Antonarakis S.E., Bonne-Tamir B.;  
 RT "Refined localization of autosomal recessive nonsyndromic deafness  
 RT DFNB10 locus using 34 novel microsatellite markers, genomic  
 RT structure, and exclusion of six known genes in the region.";  
 RL Genomics 68:22-29(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=21192304; PubMed=11279031;  
 RA Porsch-Oezcuemez M., Langmann T., Heimerl S., Borsukova H.,  
 RA Kaminski W.E., Drobnik W., Honer C., Schumacher C., Schmitz G.;  
 RT "The zinc finger protein 202 (ZNF202) is a transcriptional repressor  
 RT of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene  
 RT expression and a modulator of cellular lipid efflux.";  
 RL J. Biol. Chem. 276:12427-12433(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4; 5; 6 AND 7).  
 RX MEDLINE=21092576; PubMed=11162488;  
 RA Lorkowski S., Rust S., Engel T., Jung E., Tegelkamp K., Galinski E.A.,  
 RA Assmann G., Cullen P.;  
 RT "Genomic sequence and structure of the human ABCG1 (ABC8) gene.";  
 RL Biochem. Biophys. Res. Commun. 280:121-131(2001).  
 RN [6]  
 RP SEQUENCE OF 33-678 FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97186700; PubMed=9034316;  
 RA Croop J.M., Tiller G.E., Fletcher J.A., Lux M.L., Raab E.,  
 RA Goldenson D., Arciniegas S., Son D., Wu R.;  
 RT "Isolation and characterization of a mammalian homolog of the  
 RT Drosophila white gene.";  
 RL Gene 185:77-85(1997).  
 RN [7]  
 RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20261604; PubMed=10799558;  
 RA Venkateswaran A., Repa J.J., Lobaccaro J.-M.A., Bronson A.,  
 RA Mangelsdorf D.J., Edwards P.A.;  
 RT "Human white/murine ABC8 mRNA levels are highly induced in  
 RT lipid-loaded macrophages. A transcriptional role for specific  
 RT oxysterols.";  
 RL J. Biol. Chem. 275:14700-14707(2000).  
 RN [8]

RP INDUCTION, AND PROBABLE FUNCTION.  
 RX MEDLINE=20105556; PubMed=10639163;  
 RA Klucken J., Buechler C., Orso E., Kaminski W.E.,  
 RA Porsch-Oezcueruemez M., Liebisch G., Kapinsky M., Diederich W.,  
 RA Drobnik W., Dean M., Allikmets R., Schmitz G.;  
 RT "ABCG1 (ABC8), the human homolog of the Drosophila white gene, is a  
 RT regulator of macrophage cholesterol and phospholipid transport.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:817-822(2000).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=21474438; PubMed=11590207;  
 RA Schmitz G., Langmann T., Heimerl S.;  
 RT "Role of ABCG1 and other ABCG family members in lipid metabolism.";  
 RL J. Lipid Res. 42:1513-1520(2001).  
 CC -!- FUNCTION: Transporter involved in macrophage lipid homeostasis. Is  
 CC an active component of the macrophage lipid export complex. Could  
 CC also be involved in intracellular lipid transport processes. The  
 CC role in cellular lipid homeostasis may not be limited to  
 CC macrophages.  
 CC -!- SUBUNIT: May form heterodimers with several heterologous partners  
 CC of the ABCG subfamily.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Predominantly  
 CC localized in the intracellular compartments mainly associated with  
 CC the endoplasmic reticulum (ER) and Golgi membranes.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=7;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=1;  
 CC IsoId=P45844-1; Sequence=Displayed;  
 CC Name=2; Synonyms=J;  
 CC IsoId=P45844-2; Sequence=VSP\_000047, VSP\_000051;  
 CC Name=3; Synonyms=ABDE;  
 CC IsoId=P45844-3; Sequence=VSP\_000048, VSP\_000051;  
 CC Name=4; Synonyms=G;  
 CC IsoId=P45844-4; Sequence=VSP\_000051;  
 CC Name=5; Synonyms=F;  
 CC IsoId=P45844-5; Sequence=VSP\_000049, VSP\_000051;  
 CC Name=6; Synonyms=HI;  
 CC IsoId=P45844-6; Sequence=VSP\_000046, VSP\_000051;  
 CC Name=7; Synonyms=C;  
 CC IsoId=P45844-7; Sequence=VSP\_000050, VSP\_000051;  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN SEVERAL TISSUES.  
 CC -!- INDUCTION: Strongly induced in monocyte-derived macrophages during  
 CC cholesterol influx. Conversely, mRNA and protein expression are  
 CC suppressed by lipid efflux. Induction is mediated by the liver X  
 CC receptor/retinoide X receptor (LXR/RXR) pathway.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X91249; CAA62631.1; ALT\_INIT.  
DR EMBL; AP001746; BAA95530.1; ALT\_INIT.  
DR EMBL; AB038161; BAB13728.2; ALT\_INIT.  
DR EMBL; AJ289137; CAC00730.1; ALT\_INIT.  
DR EMBL; AJ289138; CAC00730.1; JOINED.  
DR EMBL; AJ289139; CAC00730.1; JOINED.  
DR EMBL; AJ289140; CAC00730.1; JOINED.  
DR EMBL; AJ289141; CAC00730.1; JOINED.  
DR EMBL; AJ289142; CAC00730.1; JOINED.  
DR EMBL; AJ289143; CAC00730.1; JOINED.  
DR EMBL; AJ289144; CAC00730.1; JOINED.  
DR EMBL; AJ289145; CAC00730.1; JOINED.  
DR EMBL; AJ289146; CAC00730.1; JOINED.  
DR EMBL; AJ289147; CAC00730.1; JOINED.  
DR EMBL; AJ289148; CAC00730.1; JOINED.  
DR EMBL; AJ289149; CAC00730.1; JOINED.  
DR EMBL; AJ289150; CAC00730.1; JOINED.  
DR EMBL; AJ289151; CAC00730.1; JOINED.  
DR EMBL; AF323658; AAK28836.1; -.  
DR EMBL; AF323644; AAK28836.1; JOINED.  
DR EMBL; AF323645; AAK28836.1; JOINED.  
DR EMBL; AF323646; AAK28836.1; JOINED.  
DR EMBL; AF323647; AAK28836.1; JOINED.  
DR EMBL; AF323648; AAK28836.1; JOINED.  
DR EMBL; AF323649; AAK28836.1; JOINED.  
DR EMBL; AF323650; AAK28836.1; JOINED.  
DR EMBL; AF323651; AAK28836.1; JOINED.  
DR EMBL; AF323652; AAK28836.1; JOINED.  
DR EMBL; AF323653; AAK28836.1; JOINED.  
DR EMBL; AF323654; AAK28836.1; JOINED.  
DR EMBL; AF323655; AAK28836.1; JOINED.  
DR EMBL; AF323656; AAK28836.1; JOINED.  
DR EMBL; AF323657; AAK28836.1; JOINED.  
DR EMBL; AF323664; AAK28842.1; -.  
DR EMBL; AF323658; AAK28833.1; -.  
DR EMBL; AF323640; AAK28833.1; JOINED.  
DR EMBL; AF323645; AAK28833.1; JOINED.  
DR EMBL; AF323646; AAK28833.1; JOINED.  
DR EMBL; AF323647; AAK28833.1; JOINED.  
DR EMBL; AF323648; AAK28833.1; JOINED.  
DR EMBL; AF323649; AAK28833.1; JOINED.  
DR EMBL; AF323650; AAK28833.1; JOINED.  
DR EMBL; AF323651; AAK28833.1; JOINED.  
DR EMBL; AF323652; AAK28833.1; JOINED.  
DR EMBL; AF323653; AAK28833.1; JOINED.  
DR EMBL; AF323654; AAK28833.1; JOINED.  
DR EMBL; AF323655; AAK28833.1; JOINED.  
DR EMBL; AF323656; AAK28833.1; JOINED.  
DR EMBL; AF323657; AAK28833.1; JOINED.  
DR EMBL; AF323660; AAK28838.1; -.  
DR EMBL; AF323663; AAK28841.1; ALT\_INIT.  
DR EMBL; AF323658; AAK28835.1; -.  
DR EMBL; AF323642; AAK28835.1; JOINED.  
DR EMBL; AF323645; AAK28835.1; JOINED.  
DR EMBL; AF323646; AAK28835.1; JOINED.  
DR EMBL; AF323647; AAK28835.1; JOINED.  
DR EMBL; AF323648; AAK28835.1; JOINED.

DR EMBL; AF323649; AAK28835.1; JOINED.

Query Match 17.9%; Score 596.5; DB 1; Length 678;  
Best Local Similarity 26.5%; Pred. No. 4.6e-35;  
Matches 165; Conservative 142; Mismatches 265; Indels 51; Gaps 14;

```
Qy      44 SYSVSHRVRPWWDDITSCROQWTRQILKDVSLYVESGQIMCILGSSGSGKTTLLDAMSGRL 103
      ||| |   |||   |:: : :|| :|   ||::: |:| ||:|:|:|: :||
Db      83 SYSVPE--GPWW-----RKKGYKTLLKGISGKFNSGELVAIMGPSGAGKSTLMNILAGY- 134

Qy     104 GRAGTFLGEVYVNGRALRREQFQDCFSYVLQSDTLLSSLTVRETLYHTALLAIRRGNPGS 163
      |   | | :||   |:: :|| | | | ||:| : : | | : : : |
Db     135 -RETGMKGAVLINGLPRDLRCFRKVSCYIMQDDMLLPHLTVQEAMMVAHLKLQEKDEGR 193

Qy     164 FQKKVEAVMAELSLSHVADRLIGNYSLGGISTGERRRVSIAAQLLQDPKVMFLFDEPTTGL 223
      :: |: :: | | | | | | | | | | | | | | | | | | | | |
Db     194 -REMVKEILTALGLLSCA-----NTRTGSLSGGQKRKLAIALELVNPPVMFFDEPTSGL 247

Qy     224 DCMTANQIVVLLVELARRNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDF 283
      | : | :| | | | | | | | | | | | | | | | | | | | |
Db     248 DSASCFQVVSLMKGLAQGGRSIICTIHQPSAKLFELFDQLYVLSQGCVCYRGKVCNLVPY 307

Qy     284 FNDGYPCEHSNPFDFYMDLTSVDTQSKEREIETSKRVQMIESAYKKSA-----I 334
      | | | | : | | | | | | | | | | | | | | | | | | |
Db     308 LRDLGLNCPTYHNPADFMVEVASGEYGDQNSRLVRVREGMCDSDHKRDLGGDAEVNPF 367

Qy     335 CHKTLKNIERMKHLKTLPMVPFKTKDSPGV-----FSKLGVLLRRVTRNLVRNKL 384
      | : : :: : | | | | | | | | | | | | | | | | | | |
Db     368 WHRPSEEVKQTKRLKGL-----RKDSSSMEGCHSFSASCLTQFCILFKRTFLSIMRDSV 421

Qy     385 AVITRLLQNLIMGLFLLFFVLRVRSNVLKGAIQDRVGLLYQFVGATPYTGMLNAVNLF 444
      | : : :| | | | | | | | | | | | | | | | | | | |
Db     422 LTHLRITSHIGIGLLIGLLYLIGIGNEAKK--VLSNSGFLFFSMLFLMFAALMPTVLT 479

Qy     445 LRAVSDQESQDGLYQKWQMLLAYALHVLFPFSVVATMIFSSVCYWTGLHPEVARFGYFSA 504
      | : | : | | | | | | | | | | | | | | | | | | |
Db     480 EMGVFLREHLNYWYSLKAYYLAKTMADVFPQIMFPVAYCSIVYWMTSQPSDAVRFVFAA 539

Qy     505 ALLAPHLIGEFLTLVLLGIVQNPNIIVNSVVALLSIAGVLVSGFRLNIQEMPIPKIISY 564
      | : : | | | | | | | | | | | | | | | | | | | |
Db     540 LGTMTSLVAQSLGL-LIGAASLQVATFVGVPVTAIPVLLFSGFFVSFDTIPTYLQWMSY 598

Qy     565 FTFQKYCSEILVVNEFYGLNFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPGATSRFTMNF 624
      :: :| | : :: : ||| : : : | | | : | : : : : |
Db     599 ISYVRYGFEGVILS-IYGLD----REDLHCDIDETCHF-QKSEAILRELDVENAKLYLDF 652

Qy     625 LILYSFIPALVILGIVV--FKIR 645
      ::| | :| : : | :|||
Db     653 IVLGIFFISLRLIAYFVRLRYKIR 675
```

RESULT 14

WHIT\_CERCA

ID WHIT\_CERCA STANDARD; PRT; 679 AA.

AC Q17320;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE White protein.  
 GN W.  
 OS Ceratitis capitata (Mediterranean fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Tephritoidea; Tephritidae; Ceratitis.  
 OX NCBI\_TaxID=7213;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96123276; PubMed=8533095;  
 RA Zwiebel L.J., Saccone G., Zacharopoulou A., Besansky N.J.,  
 RA Favia G., Collins F.H., Louis C., Kafatos F.C.;  
 RT "The white gene of Ceratitis capitata: a phenotypic marker for  
 RT germline transformation.";  
 RL Science 270:2005-2007(1995).  
 CC -!- FUNCTION: May be part of a membrane-spanning permease system  
 CC necessary for the transport of pigment precursors into pigment  
 CC cells responsible for eye color.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X89933; CAA61998.1; -.  
 DR InterPro; IPR003593; AAA\_ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR InterPro; IPR005284; Pigment\_permease.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR TIGRFAMs; TIGR00955; 3a01204; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS50893; ABC\_TRANSPORTER\_2; 1.  
 KW Pigment; ATP-binding; Transmembrane; Transport.  
 FT NP\_BIND 121 128 ATP (BY SIMILARITY).  
 FT TRANSMEM 427 445 POTENTIAL.  
 FT TRANSMEM 457 477 POTENTIAL.  
 FT TRANSMEM 507 525 POTENTIAL.  
 FT TRANSMEM 534 555 POTENTIAL.  
 FT TRANSMEM 568 586 POTENTIAL.  
 FT TRANSMEM 651 670 POTENTIAL.  
 FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 679 AA; 75145 MW; 3F9CBC78A835C4CC CRC64;  
  
 Query Match 17.8%; Score 591; DB 1; Length 679;  
 Best Local Similarity 28.4%; Pred. No. 1.1e-34;  
 Matches 176; Conservative 125; Mismatches 231; Indels 88; Gaps 18;  
  
 Qy 66 RQILKDVSLYVESGQIMCILGSSSGSKTTLLDAMSGRLGRAGTFLGEV--YVNGRALRRE 123

```

      : :||: |      |::: ::|||:|||||:| : | :      :|| : :
Db      101 KHLKND SGVAYPGELLAVMGSSGAGKTTLLNASAFRSSKGVQISPSTIRMLNGHPVDAK 160

Qy      124 QFQDCFSYVLQSDTLLSSLTVRETLHYTALLAIRRGNPGSFQK--KVEAVMAELSLSHVA 181
      : | :|| | | : ||| || | :|:: : | : | | :|| :|||
Db      161 EMQARCAVYQDDDLFIGSLTAREHLIFQAMVRMPR-HMTQKQKVQRVDQVIQDLSLGKCQ 219

Qy      182 DRLIG-NYSLGGISTGERRRVSIAAQLLQDPKVMLFDEPTTGLDCMTANQIVVLLVELAR 240
      : ||| : | :| |||:|:: |:: | || ::| |||:|| | : :| :| :|::
Db      220 NTLIGVPGRVKGLSGGERKRLAFASEALTDPPLLICDEPTSGLDSEMAHSVVQVLKKLSQ 279

Qy      241 RNRIVVLTIHQPRSELFQLFDKIAILSFGELIFCGTPAEMLDFFNDCGYPCPEHSNPFDF 300
      : : |:||||| |||:||||| ::| : | || | :|||: | || : | ||
Db      280 KGKTVILTIHQPSSELFELFDKILLMAEGRVAFGLGTPGEAVDFFSYIGATCPTNYTPADF 339

Qy      301 YMDLTSVDTQSKEREIETSKRVQMIESAYKKSACHKTLKNIERMKHLKTLPMVPFKTKD 360
      | : :| |||: || | : : : :| : | : | :|
Db      340 YVQVLAV---VPGREVESRDRVAKICDNFAVGKVSREMEQNFQ-----KLVKSNGFGKED 391

Qy      361 -----SPGVFSKLGVLRLRRVTRNLVRNKLAVITRLLQNLIMGLFL-LFFVLRVRSNVLK 413
      | : :| | ::| | ||| : : : | | : : |
Db      392 ENEYTYKASWFMQFRAVLWRSWLSVLKEPLLVKVRLQLTMTMAVLIGLIFLGQQLTQV-- 449

Qy      414 GAIQDRVGLLYQFVGATPYTGMLNAVNLFVLRVAVSDQESQDGLYQKWQMMLAYALHVLP 473
      : : | :| : : : :| | :|:: ||: | : ||
Db      450 -GVMNINGAIFLFLTNMTFQNSFATITVFTTELPVFMRETRSRLYRCDTYFLGKTIAELP 508

Qy      474 FSVVATMIFSSVCYWTGLGLHPEVARFGYFSAALLAPHLIGEFLTLVLLGIVQNPNIIVNSV 533
      :| :|::| :|| | | | | | | :| | :| |
Db      509 LFLVVPFLFTAIAAYPLIGLRPGVDHF-----FTALALVTLV--ANVSTSF 551

Qy      534 VALLS-----IAGVLVSGSGLRNIQEMPIPFKIISYFTFQKYCSEILVV 577
      | :| | | | | | | :| :|| :| :| :| :| :|
Db      552 GYLISCACSSSMALSVGPPVPIPFLLFGGFFLNSGSPVYFKWLSYLSWFRYANEGLLI 611

Qy      578 NEFYGL---NFTCGSSNVSVTTNPMCAFTQGIQFIEKTC--GATSRFTMNFL---ILYS 629
      |:: : || || || | | | :|| : :
Db      612 NQWADV KPG EITCTLSNT-----TCPSSGEVILETLNFSASDL PFD 652

Qy      630 FIP-ALVILGIVVFKIRDHL 648
      || ||:|:| | :| :
Db      653 FIGLALLIVG---FRISAYI 669

```

RESULT 15

ABG4\_HUMAN

ID ABG4\_HUMAN STANDARD; PRT; 646 AA.

AC Q9H172;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE ATP-binding cassette, sub-family G, member 4.

GN ABCG4 OR WHITE2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21518231; PubMed=11606068;  
RA Engel T., Lorkowski S., Lueken A., Rust S., Schlueter B., Berger G.,  
RA Cullen P., Assmann G.;  
RT "The human ABCG4 gene is regulated by oxysterols and retinoids in  
RT monocyte-derived macrophages.";  
RL Biochem. Biophys. Res. Commun. 288:483-488(2001).

RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [3]  
RP SEQUENCE OF 20-646 FROM N.A.  
RC TISSUE=Dorsal root ganglion;  
RX MEDLINE=22170423; PubMed=12183068;  
RA Oldfield S., Lowry C., Ruddick J., Lightman S.;  
RT "ABCG4: a novel human white family ABC-transporter expressed in the  
RT brain and eye.";  
RL Biochim. Biophys. Acta 1591:175-179(2002).  
CC -!- FUNCTION: May be involved in macrophage lipid homeostasis.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
CC -!- SIMILARITY: Belongs to the ABC transporter family. ABCG (White)  
CC subfamily.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AJ308237; CAC87131.1; -.  
DR EMBL; BC041091; AAH41091.1; -.  
DR EMBL; AJ300465; CAC17140.1; -.  
DR PIR; JC7777; JC7777.





Db 385 ILRDTVLTHLRFSHVIGVLIGLLYLHIGDDASK--VFNNTGCLFFSMLFLMFAALMPT 442  
 Qy 439 VNLFPLRAVSDQESQDGLYQKWQMMLAYALHVLFFSVVATMIFSSVCYWTGLGLHPEVAR 498  
 | ||: || :| :| || : :|| || :: |: || | :|  
 Db 443 VLTFPLEMAVFMREHLNYWYSLKAYYLAKTMADVFPQVVCVVCYCSIVYWMTGQPAETSR 502  
 Qy 499 FGYFSAALLAPHLIGEFLLTVLLGIVQNPINVSVALLSIAGVLVGSGFLRNIQEMPIP 558  
 | ||| | |: :| | |:| | :| :: ||: ||| : : :|  
 Db 503 FLLFSALATATALVAQSLGL-LIGAASNQLQVATFVGPGVTAIPVLLFSGFFVSFKTIPTY 561  
 Qy 559 FKIIISYFTFQKYCSEILVVNEFYGL---NFTCGSSNVSVTTNPMCAFTQGIQFIEKTCPG 615  
 : || :: :| | ::| ||: : || | | : | | :  
 Db 562 LQWSSYLSYVRYGFEGVILT-IYGMERGDLTC-----LEERCPFPREP-QSILRALDV 611  
 Qy 616 ATSRFTMNFELILYSFIPALVILGIVVFKIR 645  
 :: |:|:| | || :| :| :|  
 Db 612 EDAKLYMDFLVLGIFFLALRLLAYLVLRYS 641

Search completed: February 27, 2004, 07:12:39  
 Job time : 12.0797 secs